Boston College Law Review

Volume 60 | Issue 7 Article 6

10-30-2019

Free, Hateful, and Posted: Rethinking First Amendment Protection of Hate Speech in a Social Media World

Lauren E. Beausoleil
Boston College Law School, lauren.beausoleil@bc.edu

Follow this and additional works at: https://lawdigitalcommons.bc.edu/bclr

 Part of the First Amendment Commons, and the Internet Law Commons

Recommended Citation

Lauren E. Beausoleil, *Free, Hateful, and Posted: Rethinking First Amendment Protection of Hate Speech in a Social Media World*, 60 B.C.L. Rev. 2100 (2019), https://lawdigitalcommons.bc.edu/bclr/vol60/iss7/6

This Notes is brought to you for free and open access by the Law Journals at Digital Commons @ Boston College Law School. It has been accepted for inclusion in Boston College Law Review by an authorized editor of Digital Commons @ Boston College Law School. For more information, please contact nick.szydlowski@bc.edu.

FREE, HATEFUL, AND POSTED: RETHINKING FIRST AMENDMENT PROTECTION OF HATE SPEECH IN A SOCIAL MEDIA WORLD

Abstract: Speech is meant to be heard, and social media allows for exaggeration of that fact by providing a powerful means of dissemination of speech while also distorting one's perception of the reach and acceptance of that speech. Engagement in online "hate speech" can interact with the unique characteristics of the Internet to influence users' psychological processing in ways that promote violence and reinforce hateful sentiments. Because hate speech does not squarely fall within any of the categories excluded from First Amendment protection, the United States' stance on hate speech is unique in that it protects it. This Note argues that the harms of hate speech, when combined with the psychological impacts of social media on users, require us to accept that existing First Amendment doctrine simply is unable to accommodate the new modes of communications afforded by cyberspace and to amend the doctrine accordingly.

INTRODUCTION

On October 27, 2018, Robert Bowers, armed with three handguns and an assault rifle, entered a Pittsburgh, Pennsylvania synagogue and opened fire, murdering eleven individuals.[1] His motivation was anti-Semitism, which he had expressed on the social media website Gab prior to the killing and reaffirmed afterwards, while in custody.[2] This massacre was not an isolated incident, but

[1] Saeed Ahmed & Paul P. Murphy, *Here's What We Know So Far About Robert Bowers, the Pittsburgh Synagogue Shooting Suspect*, CNN (Oct. 28, 2018), https://www.cnn.com/2018/10/27/us/synagogue-attack-suspect-robert-bowers-profile/index.html [https://perma.cc/MY9W-2MMY]; Campbell Robertson et al., *11 Killed in Synagogue Massacre; Suspect Charged with 29 Counts*, N.Y. TIMES (Oct. 27, 2018), https://www.nytimes.com/2018/10/27/us/active-shooter-pittsburgh-synagogue-shooting.html [https://perma.cc/F8E5-CSLH].

[2] Ahmed & Murphy, *supra* note 1; Robertson et al., *supra* note 1. Prior to the killing, Bowers told a SWAT officer that he wanted all Jews dead and that "they (Jews) were committing genocide to his people." Ahmed & Murphy, *supra* note 1. Two independent reports, one from Columbia University and the other from the Anti-Defamation League, came out just before this massacre, with both reporting spikes in anti-Semitic activity online. *See* David Ingram, *Attacks on Jewish People Rising on Instagram and Twitter, Researchers Say*, NBC NEWS (Oct. 27, 2018), https://www.nbcnews.com/tech/tech-news/attacks-jewish-people-rising-instagram-twitter-researchers-say-n925086 [https://perma.cc/6KWL-C8YR] (discussing these reports). The Anti-Defamation League released its report just one day before the massacre, on October 26, 2018, and detailed the recent increase in anti-Semitic harassment online based on "analysis of computational propaganda, the Jewish American community, and the 2018 elections." CTR. ON TECH. & SOC'Y, ANTI-DEFAMATION LEAGUE, COMPUTATIONAL PROPAGANDA, JEWISH-AMERICANS

rather part of a trend of individuals committing horrific crimes after engaging in hate speech online.[3] For example, in 2015, Dylann Roof joined a bible study at the Emanuel African Methodist Church in Charleston, South Carolina—the oldest black church in the south—only to shoot and kill nine church members.[4] Prior to this mass murder, Roof had "self-radicalized," as demonstrated by a website he created and used to express his white-supremacist views.[5] In Wisconsin, in 2012, white supremacist Wade Michael Page murdered six individuals in a Sikh temple.[6] Page's online postings contained various references to Hitler, white supremacy, and even posts insisting that "passive submission [amounts to] indirect support to the oppressors" and encouraging people to stand up and

---

AND THE 2018 MIDTERMS: THE AMPLIFICATION OF ANTI-SEMITIC HARASSMENT ONLINE 6 (2018), https://www.adl.org/media/12028/download [https://perma.cc/VWU2-TW59].

[3] *See* Rachel Hatzipanagos, *How Online Hate Turns into Real-Life Violence*, WASH. POST (Nov. 30, 2018), https://www.washingtonpost.com/nation/2018/11/30/how-online-hate-speech-is-fueling-real-life-violence/ [https://perma.cc/Y6DG-USKT] (detailing an Anti-Defamation League report that stated that white supremacists committed "18 of the 34 murders documented by domestic extremists [in 2017]"). Hate speech generally refers to speech made with intention to insult, offend, demean, intimidate, or express hatred towards a person or group based on a trait, including but not limited to religion, race, national origin, sexual orientation, or disability, although precise definitions vary. *See* Matal v. Tam, 137 S. Ct. 1744, 1764 (2017) (citing United States v. Schwimmer, 279 U.S. 644, 655 (1929) (Holmes, J., dissenting)) (recognizing "speech that demeans on the basis of race, ethnicity, gender, religion, age, disability, or any other similar ground" as hateful, but constitutionally protected speech); *Speech*, BLACK'S LAW DICTIONARY (11th ed. 2019) [hereinafter *Speech*, BLACK'S LAW DICTIONARY] (defining hate speech as "[s]peech whose sole purpose is to demean people on the basis of race, ethnicity, gender, religion, age, disability, or some other similar ground, esp[ecially] when the communication is likely to provoke violence"); Rachel Weintraub-Reiter, *Hate Speech Over the Internet: A Traditional Constitutional Analysis or a New Cyber Constitution?*, 8 B.U. PUB. INT. L.J. 145, 149 (1998) (offering various definitions of hate speech). This problem is not uniquely American. *See, e.g.*, Kristen Gelineau & Jon Gambrell, *New Zealand Mosque Shooter Is a White Nationalist Who Hates Immigrants, Documents and Video Reveal*, CHI. TRIB. (Mar. 15, 2019), https://www.chicagotribune.com/nation-world/ct-mosque-killer-white-supremacy-20190315-story.html [https://perma.cc/UJ5N-5VDA] (discussing attack in New Zealand). On March 15, 2019, a 28-year-old Australian man attacked two mosques in Christchurch, New Zealand, leaving fifty-one dead and forty-nine injured. *Id.* The shooter had a social media presence, and posted a seventy-four-page manifesto filled with hate speech and white supremacist rhetoric prior to the attack. *Id.* The manifesto paid tribute to two other white nationalist attacks. *Mass Shootings: Is 'White Terrorism' Now the Main Threat in the United States?*, SOUTH CHINA MORNING POST (Aug. 5, 2019), https://www.scmp.com/news/world/united-states-canada/article/3021403/mass-shootings-white-terrorism-now-main-threat [https://perma.cc/5NUR-PJMH].

[4] Ray Sanchez & Ed Payne, *Charleston Church Shooting: Who Is Dylann Roof?*, CNN (Dec. 16, 2016), https://www.cnn.com/2015/06/19/us/charleston-church-shooting-suspect/index.html [https://perma.cc/3T7V-SX7P]; Benjy Sarlin, *Nine Dead in Charleston Church Massacre*, MSNBC (June 17, 2015), http://www.msnbc.com/msnbc/charleston-police-church-shooting [https://perma.cc/SYE7-QRJD].

[5] Hatzipanagos, *supra* note 3. On his website, Roof detailed his white supremacist views in a manifesto, criticizing blacks as inferior and criticizing others who do nothing to further white supremacy. Frances Robles, *Dylann Roof Photos and a Manifesto Are Posted on Website*, N.Y. TIMES (June 20, 2015), https://www.nytimes.com/2015/06/21/us/dylann-storm-roof-photos-website-charleston-church-shooting.html [https://perma.cc/L562-AZKL]. He complained that "[w]e have no skinheads, no real KKK, no one doing anything but talking on the internet." *Id.* After explaining that "someone has to have the bravery to take it to the real world," Roof concluded that he had "no choice" but to act. *Id.*

[6] Hatzipanagos, *supra* note 3.

"spread the truth" about white supremacy.[7] Even the pipe bomb mailer, Cesar Sayoc Jr., who allegedly sent out fourteen pipe bombs—none of which detonated—to eminent Democrats, had been posting "hateful and provocative messages" on social media websites.[8]

---

[7] Michael Laris et al., *Excessive Drinking Cost Wade Michael Page Military Career, Civilian Job*, WASH. POST (Aug. 7, 2002), https://www.washingtonpost.com/world/national-security/excessive-drinking-cost-wade-michael-page-military-career-civilian-job/2012/08/07/274ccc7a-e095-11e1-a421-8bf0f0e5aa11_story.html?utm_term=.f28fd5aa1bcb [https://perma.cc/6RM5-YMWS].

[8] Faith Karimi, *5 Days, 14 Potential Bombs and Lots of Questions. Here's What We Know*, CNN (Oct. 27, 2018), https://www.cnn.com/2018/10/26/politics/pipe-bombs-suspicious-packages-what-we-know/index.html [https://perma.cc/Z35G-FPDG]; Charlene Li, *To Head Off Mass Shootings, We Need Better Technology—Not Less*, CNN (Nov. 21, 2018), https://www.cnn.com/2018/11/21/perspectives/hate-speech-social-media-technology/index.html [https://perma.cc/QUU6-7WZR]. Sayoc was active on both Facebook and Twitter accounts, one of which he used to threaten a political analyst. Donnie O'Sullivan, *Bomb Suspect Threatened People on Twitter, and Twitter Didn't Act*, CNN (Oct. 27, 2018), https://www.cnn.com/2018/10/26/tech/cesar-sayoc-twitter-response/index.html [https://perma.cc/5MKY-SDGK]; *see also* Wesley Lowery et al., *In the United States, Right-Wing Violence Is on the Rise*, WASH. POST (Nov. 25, 2018), https://www.washingtonpost.com/national/in-the-united-states-right-wing-violence-is-on-the-rise/2018/11/25/61f7f24a-deb4-11e8-85df-7a6b4d25cfbb_story.html?utm_term=.de2a4cc08774 [https://perma.cc/V5U7-LXA5] (discussing the rise of domestic terror attacks experienced in the United States in recent years); Kevin Roose, *Cesar Sayoc's Path on Social Media: From Food Photos to Partisan Fury*, N.Y. TIMES (Oct. 27, 2018), https://www.nytimes.com/2018/10/27/technology/cesar-sayoc-facebook-twitter.html [https://perma.cc/X2UN-DCWK] (describing Cesar Sayoc's social media activity as reflecting "a fascination with Islamist terrorism, illegal immigration and anti-Clinton conspiracy theories"). By the end of the summer of 2019, in events too recent to receive the attention or analysis they deserve in this Note, the United States suffered from even more violent attacks carried out by individuals who engaged in hate speech online. *See* Elisha Fieldtadt & Ken Dilanian, *White Nationalism-Fueled Violence Is on the Rise, but FBI Is Slow to Call It Domestic Terrorism*, NBC NEWS (Aug. 5, 2019), https://www.nbcnews.com/news/us-news/white-nationalism-fueled-violence-rise-fbi-slow-call-it-domestic-n1039206 [https://perma.cc/B5XM-V6M8] (listing recent incidents). On April 28, 2019, a nineteen-year-old man opened fire at a mosque located outside of San Diego, killing one person. Shannon Van Sant, *Poway Shooting Latest in Series of Attacks on Places of Worship*, NPR (April 28, 2019), https://www.npr.org/2019/04/28/718043171/poway-shooting-latest-in-series-of-attacks-on-places-of-worship [https://perma.cc/LMH7-S7M7]. Prior to the attack, the shooter posted a document on an online message board, which was "almost identical to the one written by the Christchurch shooter" in New Zealand. Fieldtadt & Dilanian, *supra*; *see also* Jennifer Medina et al., *One Dead in Synagogue Shooting Near San Diego; Officials Call It Hate Crime*, N.Y. TIMES (Apr. 27, 2019), https://www.nytimes.com/2019/04/27/us/poway-synagogue-shooting.html [https://perma.cc/NTC2-CADR] ("The document, an anti-Semitic screed filled with racist slurs and white nationalist conspiracy theories, echoes the manifesto that was posted to 8chan by the gunman in last month's mosque slayings in Christchurch, New Zealand."). The document was filled with "anti-Semitic language and lauded white supremacy, nam[ing] the Christchurch shooter and the man accused of fatally shooting 11 people inside a Pittsburgh synagogue as inspirations for the attack." Fieldtadt & Dilanian, *supra*. Just a few months later, on July 28, 2019, a nineteen-year-old man fired into a crowd at a food festival in Gilroy, California, killing three people before killing himself. *Id.*; Minyvonne Burke, *Gilroy Garlic Festival Shooting Being Investigated as Domestic Terrorism by FBI*, NBC NEWS (Aug. 6, 2019), https://www.nbcnews.com/news/us-news/gilroy-garlic-festival-shooting-being-investigated-domestic-terrorism-fbi-n1039681 [https://perma.cc/89KV-PSDX]. Prior to the shooting, the man "left a note on Instagram instructing followers to read a nineteenth-century white nationalist book." Fieldtadt & Dilanian, *supra*. Less than one week after the Gilroy shooting, on August 3, 2019, a twenty-one-year-old man shot and killed twenty people at a Walmart in El Paso, Texas. *Id.* Prior to the shooting, the suspect apparently "posted an anti-immigrant screed on an anonymous extremist message board, citing the Christchurch, New Zealand, mosque shoot-

One year before the Pittsburgh massacre, the Supreme Court of the United States reaffirmed that hate speech, like speech generally, receives First Amendment protection.[9] This protection, however, is not absolute, and if the hate speech falls within any of the excepted categories, the constitutional protections afforded to First Amendment speech will not apply.[10] Courts typically invoke three categories to evaluate hate speech: incitement to imminent lawless action, fighting words, and true threats.[11] Only if the hate speech fits within one of these categories will its proscription be proper.[12] This broad protection is unique to the United States.[13] Various European countries, including Germany, ban hate speech in some form, with differences among countries reflecting the lack of consensus on a universal definition of hate speech.[14] Part I of this Note describes

---

er who left fifty-one dead in March, as an inspiration." *Id.* On August 9, 2019, police arrested a twenty-three-year-old security guard in his home in Las Vegas, where authorities found "an AR-15 rifle, bolt action rifle, bomb-making materials, and a journal in his room with a hand-drawn picture of an attack on a Las Vegas bar that he thought was frequented by gay people." Danika Fears, *Las Vegas White Supremacist Conor Climo Arrested After Threatening to Attack Synagogue, LGBTQ Bar: DOJ*, DAILY BEAST (Aug. 9, 2019), https://www.thedailybeast.com/las-vegas-white-supremacist-conor-climo-arrested-after-threatening-to-attack-synagogue-lgbtq-bar-doj [https://perma.cc/N4M3-L7Z3] (reporting arrest). The man admitted that "he began communicating with members of the neo-Nazi group the Feuerkrieg Division . . . at the end of 2017," further demonstrating this phenomenon. *Id.*

[9] *See Matal*, 137 S. Ct. at 1751 (explaining that the First Amendment protects even speech that offends).

[10] *See* Virginia v. Black, 538 U.S. 343, 358 (2003) (explaining that the state may constitutionally regulate certain categories of speech); R.A.V. v. City of St. Paul, 505 U.S. 377, 382–83 (1992) (explaining the rationale for excluding categories of speech from First Amendment protection).

[11] *See, e.g.*, NAACP v. Claiborne Hardware, 458 U.S. 886, 928 (1982) (evaluating racially charged and threatening statements under the incitement to imminent lawless action standard); Watts v. United States, 394 U.S. 705, 706, 708 (1969) (evaluating a threat to end the U.S. President's life under the true threat framework); United States v. White, 670 F.3d 498, 513 (4th Cir. 2012) (evaluating a statement calling someone "an enemy, not just to the white race but of all humanity [. . . who] must be killed" under the true threat doctrine); *In re* John M., 36 P.3d 772, 776 (Ariz. Ct. App. 2001) (evaluating a juvenile's racial slurs at African American women under the fighting words framework).

[12] Planned Parenthood of Columbia v. Am. Coal. of Life Activists, 290 F.3d 1058, 1092 (9th Cir. 2002), as amended (July 10, 2002) (Reinhardt, J., dissenting) ("Speech . . . may not be punished or enjoined unless it falls into one of the narrow categories of unprotected speech recognized by the Supreme Court: true threat, incitement conspiracy to commit criminal acts, fighting words.") (citations omitted).

[13] *See* Robert A. Khan, *Why Do Europeans Ban Hate Speech? A Debate Between Karl Loewenstein and Robert Post*, 41 HOFSTRA L. REV. 545 (2013) (exploring the sociological and historical reasons behind differences in treatment of hate speech laws in the United States and in various European countries); Mike Gonzalez, *Europe's War on Free Speech*, HERITAGE FOUND. (Feb. 9, 2018), https://www.heritage.org/europe/commentary/europes-war-free-speech [https://perma.cc/NNM3-CU3S] (discussing the law governing hate speech in the United States and comparing it with that in Poland, France, Germany, and the United Kingdom); *see also* ARTICLE 19, RESPONDING TO 'HATE SPEECH': COMPARATIVE OVERVIEW OF SIX EU COUNTRIES (2018) (providing a comparative overview of six European Union countries' treatment of hate speech).

[14] Gonzalez, *supra* note 13; Mark Scott & Janosch Delcker, *Free Speech vs. Censorship in Germany*, POLITICO (Jan. 6, 2018), https://www.politico.eu/article/germany-hate-speech-netzdg-facebook-youtube-google-twitter-free-speech/ [https://perma.cc/2JY9-FSH4]; *see* EMORE, AN OVERVIEW ON HATE CRIME AND HATE SPEECH IN 9 EU COUNTRIES 8 (2017) ("From a legal point of view, most countries do not maintain a clear definition of either hate speech or hate crime."); *see also* Alexander Tsesis,

the impact of social media on the manifestation of hate speech, along with the general psycho-sociological processes responsible for this impact.[15] Part II discusses the present First Amendment protection of hate speech in the United States.[16] Part III of this Note explores the various positions taken by participants of the debate surrounding the protection of hate speech.[17] Finally, Part IV rejects the arguments made by those in favor of protecting hate speech and argues for a change in the standard applied in evaluating hate speech, calling for reduced protection of such speech in light of the impact of technological advances on information dissemination and reception.[18]

I. PSYCHOSOCIAL IMPACTS OF SOCIAL MEDIA

Social media and related online communications may have profound effects on human psychology and social behavior.[19] These effects, and the dangers they pose, are even more pronounced when those communications involve hate speech.[20] Section A of this Part discusses hate speech, both generally and in the context of social media.[21] Section B of this Part discusses the polarization mechanisms of social media and the social and psychological impacts of such polari-

---

Hate in Cyberspace: Regulating Hate Speech on the Internet, 38 SAN DIEGO L. REV. 817, 858 (2001) (listing countries that have laws proscribing hate speech). Germany recognizes a constitutional right to freedom of expression, embodied in Article 5 of Germany's Basic Law, but this right is subject to a limitation recognizing "the citizen's right to personal respect." *Id.* at 861–62. Germany has passed various laws that can be used to penalize persons who use the Internet to share hateful messages about outgroups, including one that subjects to imprisonment those who incite people to hate certain subsets of the population, advocate violence or "arbitrary measures" against them, or slander them. *Id.* at 862. German law also holds public distribution or supply of "writings that incite to race hatred or describe cruel or otherwise inhuman acts of violence against humans in a manner which glorifies or minimizes such acts of violence or represents the cruel or inhuman aspects of the occurrence in a manner offending human dignity." *Id.* (citing Eric Stein, *History Against Free Speech: The New German Law Against the "Auschwitz"—and Other—"Lies,"* 85 MICH. L. REV. 277, 322–23 (1986) (quoting STRAFGESETZBUCH [StGB] [PENAL CODE] art. 131)). More recently, Germany passed legislation tailored towards hate speech on social media, known as NetzDG. *Germany Starts Enforcing Hate Speech Law*, BBC NEWS (Jan. 1, 2018), https://www.bbc.com/news/technology-42510868 [https://perma.cc/WT58-8DWM]. This law requires social media networks to remove "hate speech, fake news and illegal material" within twenty-four hours after such content appears online. *Id.*

[15] *See infra* notes 19–64 and accompanying text.

[16] *See infra* notes 65–126 and accompanying text.

[17] *See infra* notes 127–205 and accompanying text.

[18] *See infra* notes 206–262 and accompanying text.

[19] Bernard J. Luskin, *Brain, Behavior, and Media*, PSYCHOL. TODAY (Mar. 29, 2012), https://www.psychologytoday.com/us/blog/the-media-psychology-effect/201203/brain-behavior-and-media [https://perma.cc/KJ3V-AX92].

[20] *See, e.g.*, Tom Jacobs, *How Hate Speech Boosts Bigotry and Intolerance*, PAC. STANDARD (Dec. 4, 2017), https://psmag.com/social-justice/how-hate-speech-boosts-bigotry-and-intolerance [https://perma.cc/PM9Z-Z9QL] (discussing results of a study conducted in Poland linking hate speech with desensitization to violence).

[21] *See infra* notes 24–30 and accompanying text.

zation on social behavior, as demonstrated by academic research and studies.[22] Finally, Section C evaluates the amplification of dangers relating to these social and psychological impacts, as facilitated by social media.[23]

A. Online "Hate Speech"

Although there is no universal definition of hate speech, a standard definition of hate speech is speech made solely to express hatred toward, insult, offend, demean, or intimidate a person or group on the basis of a trait, such as religion, race, national origin, sexual orientation, or disability.[24] Online hate speech has become increasingly prevalent, and its harms have become increasingly clear.[25] Even social media companies have taken notice of this trend and have adjusted their policies accordingly.[26] Twitter, for example, released a statement explaining that the platform decided to amend its hateful conduct policy to proscribe content that dehumanizes members of a discernable group based on group membership, even in the absence of a specifically targeted individual.[27] To justify its decision, Twitter referred readers to scholars who have recognized the link

---

[22] *See infra* notes 31–48 and accompanying text.

[23] *See infra* notes 49–64 and accompanying text.

[24] *See Matal*, 137 S. Ct. at 1764 (recognizing that "speech that demeans on the basis of race, ethnicity, gender, religion, age, disability, or any other similar ground is hateful," but remains constitutionally protected); *Speech*, BLACK'S LAW DICTIONARY, *supra* note 3 (defining hate speech as "[s]peech whose sole purpose is to demean people on the basis of race, ethnicity, gender, religion, age, disability, or some other similar ground, esp[ecially] when the communication is likely to provoke violence"); Weintraub-Reiter, *supra* note 3, at 149 (offering various definitions of hate speech).

[25] *See* Yulia A. Timofeeva, *Hate Speech Online: Restricted or Protected? Comparison of Regulations in the United States and Germany*, 12 J. TRANSNAT'L L. & POL'Y 253, 256–57 (2003) ("The Internet has provided unique resources for expanding hate propaganda."); Weintraub-Reiter, *supra* note 3, at 146 ("Cyberspace has been and continues to be used to perpetuate hate speech."). Notably, those who participate in organized hate groups have used the Internet to spread their views. Weintraub-Reiter, *supra* note 3, at 148.

[26] *See* David Goldman, *Big Tech Has Made the Social Media Mess. It Has to Fix It.*, CNN BUS. (Oct. 29, 2018, 3:17 PM), https://www.cnn.com/2018/10/29/tech/social-media-hate-speech/index.html [https://perma.cc/89SG-56YH] (discussing policy changes made by Facebook, YouTube, and Twitter following incidents involving promoting hate groups through their platforms). After this Note was submitted for publication, YouTube again changed its hate speech policies—this time, taking more aggressive actions. *See* YOUTUBE, *Our Ongoing Work to Tackle Hate*, YOUTUBE: OFFICIAL BLOG (June 5, 2019), https://youtube.googleblog.com/2019/06/our-ongoing-work-to-tackle-hate.html [https://perma.cc/4V43-T3LX] (discussing current and past changes in policy). The company announced:

> Today, we're taking another step in our hate speech policy by specifically prohibiting videos alleging that a group is superior in order to justify discrimination, segregation or exclusion based on qualities like age, gender, race, caste, religion, sexual orientation or veteran status. This would include, for example, videos that promote or glorify Nazi ideology, which is inherently discriminatory.

Id.

[27] Vijaya Gadde & Del Harvey, *Creating New Policies Together*, TWITTER: COMPANY (Sept. 28, 2018), https://blog.twitter.com/official/en_us/topics/company/2018/Creating-new-policies-together.html [https://perma.cc/5PAQ-56XG].

between dehumanizing language and increased violence.[28] Reports on the increasing prominence of hate speech online have led other nations, and advocates of restrictions on hate speech, to call for action.[29] Similarly, statistics on the rise of hate crimes committed in the United States in recent years, which show a sharp increase of seventeen percent in 2017 alone, have led to concerns about the implications of online hate speech.[30]

B. The Creation of Echo Chambers

Social media acts as a polarization medium, both through users' affirmative actions and less-deliberate mechanisms relating to how one engages and interacts on social media.[31] Social media provides users with individually tailored receipts of information, which users control through their selections of friends, followers, accounts followed, and other interactions online.[32] Studies suggest

---

[28] *Id.* (first citing DANGEROUS SPEECH PROJECT, *What Is Dangerous Speech?*, https://dangerousspeech.org/the-dangerous-speech-project-preventing-mass-violence/ [https://perma.cc/88ZB-ZM2X] (discussing Susan Benesch's research); then citing Herbert C. Kelman, *Violence Without Moral Restraint: Reflections on the Dehumanization of Victims and Victimizers*, 29 J. SOC. ISSUES 25, 25–61 (1973)). Susan Benesch was noted for her work recognizing dangerous speech and its role in dehumanization and normalizing violence, while Herbert C. Kelman suggested that such dehumanization may reduce the ability for moral judgments to resist and combat unjustifiable violence. *See id.* (explaining the rationale for policy changes).

[29] *See, e.g.*, Tom Batchelor, *Neo-Nazis Benefiting from Dramatic Rise in Racist Websites to Spread Hate and Incite Violence, UN Warns*, INDEPENDENT (Nov. 1, 2018), https://www.independent.co.uk/news/world/neo-nazi-racism-far-right-social-media-hate-speech-twitter-facebook-un-a8613496.html [https://perma.cc/5VSP-HM3U] (relaying a United Nations expert's warning of increases in attacks on various minority groups and communities and noting that the United Nations' special rapporteur for racism has reported a 600% increase since 2012 in the number of individuals expressing white-supremacist views on Twitter); IRISH HUM. RTS. & EQUALITY COMMISSION, *Press Release: Human Rights and Equality Commission Challenges Rise of Hate Speech Online* (Nov. 28, 2018), https://www.ihrec.ie/human-rights-and-equality-commission-challenges-rise-of-hate-speech-online/ [https://perma.cc/GQ36-MB53] (discussing hate speech's increasingly prominent presence online and urging for Ireland to take on the role of an international leader in fighting the spread of hate speech on the Internet).

[30] *See* Brian Levin et al., *New Data Shows U.S. Hate Crimes Continued to Rise in 2017*, THE CONVERSATION (June 26, 2019, 6:41 AM), http://theconversation.com/new-data-shows-us-hate-crimes-continued-to-rise-in-2017-97989 [https://perma.cc/B3UH-HWXC] (discussing trends in hate crime statistics since 1992); FBI, *2017 Hate Crime Statistics Released* (Nov. 13, 2018), https://www.fbi.gov/news/stories/2017-hate-crime-statistics-released-111318 [https://perma.cc/EBE8-9XNG] (reporting rises in hate crimes statistics).

[31] *See* CASS R. SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA 1, 128–30 (2018) (discussing the role of social media in contributing to polarization and in influencing perception and behavior).

[32] *See, e.g.*, *How Do I Hide a Post That Appears in My News Feed?*, FACEBOOK, https://www.facebook.com/help/268028706671439?helpref=related [https://perma.cc/Y2K4-8NTL] (instructing on hiding content from one's Facebook newsfeed); *How to Block Accounts on Twitter*, TWITTER, https://help.twitter.com/en/using-twitter/blocking-and-unblocking-accounts [https://perma.cc/KA8K-FWYD] (instructing on blocking Twitter accounts to reduce interactions); *Using Twitter*, TWITTER, https://help.twitter.com/en/using-twitter [https://perma.cc/CZ9S-UY88]; *see also* SUNSTEIN, *supra* note 31, at 123

that another subconscious mechanism contributes to polarization on social media as well: confirmation bias.[33] A study examining the behavior of over seven hundred individuals during a six-week period, in which researchers found that people are more likely to click on links with information supporting their own views, demonstrates this phenomenon.[34] This idea of a confirmation bias—meaning that individuals are more likely to visit websites depicting material confirming, rather than undermining, their own beliefs—has also been studied and confirmed in the context of Facebook and other social media outlets.[35] These studies demonstrate "the echo chamber effect" of social media, which acts to limit the range of information that social media users absorb.[36]

In addition to muting an opposite ideological viewpoint, the creation of echo chambers can also amplify a user's own viewpoint through social reinforcement, which was first recognized by researchers studying the effect of praise on children.[37] In 1968, an experiment involving children who had spent

---

(discussing Facebook's algorithm for controlling information feeds and the fact that any information one encounters online is not random).

[33] *See* R. Kelly Garrett, *Echo Chambers Online?: Politically Motivated Selective Exposure Among Internet News Users*, 14 J. COMPUTER-MEDIATED COMM. 265, 279 (2009) (discussing the tendency of individuals to seek information that supports their own); Alexandra Andorfer, Note, *Spreading Like Wildfire: Solutions for Abating the Fake News Problem on Social Media Via Technology Controls and Government Regulation*, 69 HASTINGS L.J. 1409, 1417 (2018) (describing confirmation bias and its role in promoting the spread of fake news).

[34] Garrett, *supra* note 33, at 269–70, 279 (explaining the study, methodology, and results).

[35] SUNSTEIN, *supra* note 31, at 118–21, 123 (discussing studies evidencing the effect of confirmation bias on Facebook and homophily on Twitter); Victoria Ward, *Facebook Makes Us More Narrow Minded, Study Finds*, TELEGRAPH (Jan. 7, 2018), https://www.telegraph.co.uk/news/newstopics/howaboutthat/12086281/Facebook-makes-us-more-narrow-minded-study-finds.html [https://perma.cc/VG74-G885]; *see also* Aaron Retica, *Homophily*, N.Y. TIMES MAG. (Dec. 10, 2006), https://www.nytimes.com/2006/12/10/magazine/10Section2a.t-4.html [https://perma.cc/5XVN-LKWK] (discussing sociologists' use of the term "homophily" to "explain our inexorable tendency to link up with one another in ways that confirm rather than test our core beliefs"). For example, Itai Himelboim and his co-authors studied homophily on Twitter, finding that "users are unlikely to be exposed to cross-ideological content" from the users they followed. SUNSTEIN, *supra* note 31, at 118–19 (citing Itai Himelboim et al., *Valence-Based Homophily on Twitter: Network Analysis of Emotions and Political Talk in the 2012 Presidential Election*, 18 NEW MEDIA & SOC'Y 1382 (2014)). M.D. Conover conducted a study investigating political communication networks on Twitter, focusing on data from the 2010 midterm elections, and found an "extremely limited connectivity" between conservative and liberal users. SUNSTEIN, *supra* note 31, at 119. A 2012 study focusing on politically engaged Twitter users similarly found that users received disproportionate exposure to ideologically similar tweets. *Id.* at 120.

[36] *See* SUNSTEIN, *supra* note 31, at 114 (asserting that there is evidence of the "echo chamber effect" and providing an exemplary study demonstrating); Himelboim et al., *supra* note 35, at 1395 (concluding that the findings in a study of homophily on Twitter evidence the formation of "affirmative echo chambers," where users interact with others who share similar ideologies and expose themselves primarily to content of similar tones); Andrew Hutchinson, *Politics, Fatigue, and the Social Echo-Chamber Effect*, SOC. MEDIA TODAY (Oct. 26, 2016), https://www.socialmediatoday.com/social-networks/politics-fatigue-and-social-echo-chamber-effect [https://perma.cc/8TH9-6RWV] (discussing the echo chamber effect on social media and role in narrowing the scope of information received by users).

[37] Kendra Cherry, *Social Reinforcement and Behavior*, VERY WELL MIND (Nov. 3, 2018), https://www.verywellmind.com/what-is-social-reinforcement-2795881 [https://perma.cc/XR6Y-CDJF] (citing

little time studying was conducted to evaluate the effects of teacher attention on study behavior.[38] The researchers found that when teachers ignored non-study behavior and gave the children praise following study behavior, the children's study rates drastically increased, sometimes almost doubling.[39] The social reinforcement that the children received impacted both the children's perception of studying and actions.[40]

Within social media echo chambers, social reinforcement can act to reinforce specific arguments or viewpoints and may even encourage users to share more controversial ideas than they would have had there been a greater anticipation of other users challenging those ideas.[41] Both results relate to humans' socio-psychological tendency to seek peer approval.[42] When individuals express an opinion, they are typically sensitive to their peers' reactions—specifically, whether their peers respond with approval or disapproval.[43] Disagreement tends to decrease one's attachment to a specific idea, and approval or agreement tends to reinforce one's attachment to that expressed idea.[44] This process, which involves less explicit social reinforcement, has been connected to increases in polarization of one's viewpoint.[45] Similar to the effect on the children in the 1968 study, this social reinforcement and polarization can also lead to action.[46]

In a recent study focused on a new German far-right political group and, in particular, Facebook users engaged in hate speech on the group's page, researchers confirmed that the propagation of radical or extreme perspectives on social media contributes to polarization.[47] More importantly, however, in addition to finding that hate speech and related sentiments are propagated by social media networks, the researchers were able to identify a link between momentary

---

R. Vance Hall et al., *Effects of Teacher Attention on Study Behavior*, 1968 J. APPLIED BEHAV. ANALYSIS 1, 1); Jerry Daykin, *Could Social Media Be Tearing Us Apart?*, THE GUARDIAN (June 28, 2016), https://www.theguardian.com/media-network/2016/jun/28/social-media-networks-filter-bubbles [https://perma.cc/52ZW-2CRM].

[38] *See* Hall et al., *supra* note 37, at 1 (providing an overview of the study).

[39] *See id.* at 1, 6 (explaining results generally and evaluating increase in a student's mean study rate from thirty-seven percent to seventy-one percent).

[40] *See id.* at 10–12 (discussing results).

[41] Daykin, *supra* note 37.

[42] *See* S. Banisch & E. Olbrich, *Opinion Polarization by Learning from Social Feedback*, J. MATHEMATICAL SOC. 1, 2 (Oct. 10, 2018), https://www.tandfonline.com/doi/full/10.1080/0022250X.2018.1517761 [https://perma.cc/YN9W-CNAQ] (citing GEORGE C. HOMANS, SOCIAL BEHAVIOR: ITS ELEMENTARY FORMS (rev. ed. 1974)) (detailing the influence of peer approval in opinion generation).

[43] Banisch & Olbrich, *supra* note 42, at 2.

[44] *Id.*

[45] *Id.*

[46] *See* Hall et al., *supra* note 37, at 1 (studying the effect of praise on children); Karsten Müller & Carlo Schwarz, Fanning Flames of Hate: Social Media and Hate Crime 1, 33 (Nov. 30, 2018) (unpublished manuscript), https://ssrn.com/abstract=3082972 (linking bursts of anti-refugee sentiment to increased incidences of violence).

[47] Müller & Schwarz, *supra* note 46, at 1, 6.

"bursts" of anti-refugee sentiment in social media posts and incidents of violence targeted at refugees, thus demonstrating the real-life dangers of online hate speech.[48]

C. Amplified Echoes and Harms Online

In the context of social media, the echo chamber effect proves especially dangerous.[49] Terrorist organizations, recognizing this, have capitalized on the inherent power of social media echo chambers in order to create an online influence that appears larger than it actually is.[50] ISIS, for example, has strategically used social media to attract and radicalize individuals from all over the globe, taking advantage of social media algorithms controlling content-filtering and distribution and the echo chamber effect as amplified by social media to do so.[51] ISIS takes advantage of algorithms and 'bots' to create an apparent following far more prevalent than its actual following.[52] This assists the terrorist group in creating "echo chambers, in which all moderating influences are removed and violent voices are amplified," thus causing normalization of extremist views and violence, in particular.[53] By creating sub-communities and insulating potential recruits from moderating voices, ISIS creates a deceptive and distorted reality for individuals susceptible to their influence who then become subjected to the normalization of extremist views.[54] These efforts have not gone unrewarded, as the number of Americans charged for crimes relating to the Islamic State has steadily increased.[55]

---

[48] *Id.* at 1, 33.

[49] *See* Donna Farag, Note, *From Tweeter to Terrorist: Combatting Online Propaganda When Jihad Goes Viral*, 54 AM. CRIM. L. REV. 843, 845, 863 (2017) (discussing ISIS's successful and strategic use of social media echo chambers to radicalize individuals online).

[50] *Id.* at 856, 863.

[51] *See id.* at 855, 863 (discussing ISIS's social media tactics and strategies for recruitment and describing how online echo chambers supplement these efforts).

[52] *Id.* at 855–56 (discussing ISIS's use of social media bots and algorithms to spread its message). "Bots" refer to software that can generate social media posts that appear to, but do not, come from a human user. *Id.* at 856.

[53] *Id.* at 863 (citing Peter R. Neumann, *Options and Strategies for Countering Online Radicalization in the United States*, 36 STUD. CONFLICT & TERRORISM 431, 436 (2013)); *see also* Jaime M. Freilich, *Section 230's Liability Shield in the Age of Online Terrorist Recruitment*, 83 BROOK. L. REV. 675, 692–93 (2018) (discussing "radicalization echo chambers" and the ability of terrorists to utilize them without physically entering the United States).

[54] Farag, *supra* note 49, at 863 (citing Neumann, *supra* note 53, at 435–36). For a detailed overview of this process, along with a firsthand account of the progression of this distortion of reality by a seventeen-year-old who fell victim to ISIS recruitment tactics, see Scott Shane et al., *Americans Attracted to ISIS Find an 'Echo Chamber' on Social Media*, N.Y. TIMES (Dec. 8, 2015), https://www.nytimes.com/2015/12/09/us/americans-attracted-to-isis-find-an-echo-chamber-on-social-media.html [https://perma.cc/P348-D7P8].

[55] *See* Farag, *supra* note 49, at 844 (discussing the rising number of Americans prosecuted for charges related to the Islamic State and noting that among these individuals, a commonality was that "all, or nearly all, had spent hours on the Internet trumpeting their feelings about the Islamic State") (quoting

Even outside of terrorism, the dangers inherent in social media's echo chambers are clear.[56] The term "fake news" has garnered much attention since the 2016 presidential election campaign, as Russian forces have been accused of influencing the outcome of the election by spreading propaganda and fictitious news stories to damage Hillary Clinton's reputation and promote Donald Trump's campaign.[57] Following the election, in 2017, Congress questioned social media and technology industry giants, including Facebook, Twitter, and Google, for failing to stop the spread of such stories, with senators criticizing the companies' "limp response" to combating the problem.[58] While dubious news stories may seem innocuous enough, the ease of generating and disseminating such stories and the relatively high degree of acceptance of news stories read online as true makes the phenomenon of fake news particularly powerful.[59] "Fake news" has even led to violence by those acting on it.[60]

Adam Goldman et al., *The Islamic State's Suspected Inroads into America*, WASH. POST (Feb. 22, 2017), https://www.washingtonpost.com/graphics/national/isis-suspects/ [https://perma.cc/C52K-8NAR]). As of July, 2018, 125 Americans have been charged and, of those, seventy-six have been convicted. Goldman et al., *supra*. Social media has been a commonality among those cases. Farag, *supra* note 49, at 844; *see also* Eric Posner, *ISIS Gives Us No Choice but to Consider Limits on Speech*, SLATE (Dec. 15, 2015), http://www.slate.com/articles/news_and_politics/view_from_chicago/2015/12/isis_s_online_radicalization_efforts_present_an_unprecedented_danger.html [https://perma.cc/U2AE-KTJ4] (discussing "radicalization echo chambers" and concluding that "the change in technology, more than the change in the nature of foreign threats, has given rise to a historic and unprecedented danger from foreign radicalization and recruitment").

[56] *See* Joseph Thai, *The Right to Receive Foreign Speech*, 71 OKLA. L. REV. 269, 310–11 (2018) (suggesting the potential inability of a remedy to counter social media's polarized echo chambers); Andorfer, *supra* note 33, at 1423 (describing harms associated with fake news).

[57] *See How Russian Twitter Bots Put Out Fake News During the 2016 Election*, NPR (April 3, 2017, 4:53 PM), https://www.npr.org/sections/alltechconsidered/2017/04/03/522503844/how-russian-twitter-bots-pumped-out-fake-news-during-the-2016-election [https://perma.cc/8255-PSP8] (discussing Russians' use of bots to influence 2016 election); Tina Nguyen, *Did Russian Agents Influence the U.S. Election with Fake News?*, VANITY FAIR (Nov. 25, 2016), https://www.vanityfair.com/news/2016/11/fake-news-russia-donald-trump [perma.cc/9QKM-BECL] (discussing Russians' use of "fake news" to influence campaign); Scott Shane, *The Fake Americans Russia Created to Influence the Election*, N.Y. TIMES (Sept. 7, 2017), https://www.nytimes.com/2017/09/07/us/politics/russia-facebook-twitter-election.html [https://perma.cc/6WF2-NPTU] (discussing Russians' use of fake social media accounts to influence campaign and spread anti-Clinton messages).

[58] Andorfer, *supra* note 33, at 1411; Hamza Shaban et al., *Facebook, Google and Twitter Testified on Capitol Hill. Here's What They Said.*, WASH. POST (Oct. 31, 2017), https://www.washingtonpost.com/news/the-switch/wp/2017/10/31/facebook-google-and-twitter-are-set-to-testify-on-capitol-hill-heres-what-to-expect/ [https://perma.cc/A23S-J2BH].

[59] *See* Andorfer, *supra* note 33, at 1417 (discussing the ease with which individuals accept fake news as true and social media's facilitation of such acceptance).

[60] *See id.* at 1412 (providing an example of a violent response to "fake news"). In 2016, Edgar Welch fired an assault rifle into and subsequently searched a Washington, D.C. pizza restaurant for child sex-slaves, after reading a false news report that accused the restaurant of harboring them as part of a child-abuse ring led by Hillary Clinton. *Id.* The story was based on correspondence between the chairman of Clinton's campaign and restaurant owners in leaked emails and gained traction after its dissemination on social media, with articles and reports using the term #PizzaGate. Emma M. Savino, Comment, *Fake News: No One Is Liable, and That Is a Problem*, 65 BUFF. L. REV. 1101, 1108–09 (2017) (discussing the

The echo chamber effect is a driving force for the phenomenon of fake news, aiding in its spread, amplifying its impact, and reducing the ability of legitimate news stories to counter and dispel those false stories.[61] As these examples demonstrate, social media has become a powerful and dangerous propeller of the echo chamber effect.[62] This becomes problematic when equally powerful and dangerous forces manifest themselves through social media.[63] This is especially true where those forces receive constitutional protection as speech.[64]

II. THE FIRST AMENDMENT'S POSITION ON HATE SPEECH

The First Amendment, incorporated to the states by the Fourteenth Amendment, promises protection of speech.[65] This protection, however, is subject to various carve outs in the form of categories of speech excepted from its protection.[66] Of these categories, three—incitement to imminent lawless action, true threats, and fighting words—can apply to hate speech.[67] Section A of this Part discusses First Amendment protection of speech generally.[68] Section B of this Part discusses those three relevant categories of excepted speech.[69]

---

spread of #PizzaGate story, which began with tying the term "cheese pizza" to "child pornography" and later developed into stories involving "Satanism, kill rooms, underground tunnels, and cannibalism in other nearby businesses"); Marc Fisher et al., *Pizzagate: From Rumor, to Hashtag, to Gunfire in D.C.*, WASH. POST (Dec. 6, 2016), https://www.washingtonpost.com/local/pizzagate-from-rumor-to-hashtag-to-gunfire-in-dc/2016/12/06/4c7def50-bbd4-11e6-94ac-3d324840106c_story.html [https://perma.cc/X964-K85H] (describing the incident).

[61] *See* Thai, *supra* note 56, at 310–11 (discussing the echo chamber effect's role in social media and the European Research Council's conclusion crediting Facebook as the most significant tool in the spread of fake news); Eric Emanuelson, Jr., Comment, *Fake Left, Fake Right: Promoting an Informed Public in the Era of Alternative Facts*, 70 ADMIN. L. REV. 209, 216–17 (2018) (explaining how social media "echo chambers" intensify the problem of fake news by creating "homogenous news feeds" that reduce a user's ability to recognize misleading content).

[62] *See* Freilich, *supra* note 53, at 692–93 (identifying the dangers of "radicalization echo chambers" created by terrorist organizations in recruitment attempts); Emanuelson, supra note 61, at 216–17 (identifying the dangers of "fake news" as intensified by echo chambers); Fisher et al., *supra* note 60 (discussing the #PizzaGate story).

[63] *See, e.g.*, Fisher et al., *supra* note 60 (describing a gunman acting on the #PizzaGate story and the story's origin in speech on social media); *see also* Thai, *supra* note 56, at 310–11 (discussing the dangers and ramifications of fake news online).

[64] *See* Am. Freedom Def. Initiative v. Wash. Metro. Area Transit Auth., 898 F. Supp. 2d 73, 79 (D.D.C. 2012) (explaining that hate speech can receive First Amendment protections).

[65] Reed v. Gilbert, 135 S. Ct. 2218, 2226 (2015) (quoting U.S. CONST. amend. IV).

[66] *See* Snyder v. Phelps, 580 F.3d 206, 214 (4th Cir. 2009), *aff'd*, 562 U.S. 443 (2011) (explaining that speech may receive differing degrees of First Amendment protection). These excepted categories include fighting words, incitement to imminent lawless action, true threats, obscenity, child pornography, defamation, and speech integral to criminal conduct. DAVID L. HUDSON, JR., LEGAL ALMANAC: THE FIRST AMENDMENT: FREEDOM OF SPEECH § 2:5 (2012).

[67] *See* Virginia v. Black, 538 U.S. 343, 359 (2003) (discussing exclusion of fighting words, incitement to imminent lawless action, and true threats).

[68] *See infra* notes 70–83 and accompanying text.

[69] *See infra* notes 84–126 and accompanying text.

A. First Amendment Protection

The First Amendment protects speech and other expressive conduct from government interference, as the government generally may only regulate such conduct without favoring or disfavoring a viewpoint—the principle known as viewpoint neutrality.[70] This protection extends even to speech that expresses ideas that most people would find distasteful, offensive, disagreeable, or discomforting, and thus extends even to hate speech.[71] The United States Supreme Court has described this concept—that the government cannot proscribe speech simply because it expresses offensive ideas—as a bedrock principle of the First Amendment.[72]

First Amendment protection, however, is still not absolute, as the government can place restrictions on the time, place, and manner of speech, so long as such restrictions are reasonable, narrowly tailored, and balance the interests of all involved.[73] The context, content, and form of the speech can support lesser or greater protection.[74] Specifically, if these factors indicate that the speech ad-

---

[70] *See* Matal v. Tam, 137 S. Ct. 1744, 1766 (2017) ("The First Amendment's viewpoint neutrality principle protects more than the right to identify with a particular side. It protects the right to create and present arguments for particular positions in particular ways, as the speaker chooses."); R.A.V. v. City of St. Paul, 505 U.S. 377, 382 (1992) (discussing the First Amendment's protection against government regulation of speech based on the expressive content). Relatedly, content-based regulations, which violate the principle of content neutrality, are presumptively invalid. *R.A.V.*, 505 U.S. at 382.

[71] *Black*, 538 U.S. at 358; *see also* Am. Freedom Def. Initiative v. Wash. Metro. Area Transit Auth., 898 F. Supp. 2d 73, 79 (D.D.C. 2012) (explaining that hate speech can receive First Amendment protections).

[72] *Matal*, 137 S. Ct. at 1751; Am. Booksellers Ass'n, Inc. v. Hudnut, 771 F.2d 323, 329 (7th Cir. 1985), *aff'd*, 475 U.S. 1001 (1986) ("Most governments of the world act on this empirical regularity, suppressing critical speech. In the United States, however, the strength of the support for this belief is irrelevant. Seditious libel is protected speech unless the danger is not only grave but also imminent.").

[73] *Black*, 538 U.S. at 358; *Snyder*, 580 F.3d at 214. Where a government action or regulation impedes on an individual's constitutional rights, such as those listed in the First Amendment, it intrudes on an individual's fundamental liberty interests and therefore triggers a strict scrutiny level of judicial review. *Reed*, 135 S. Ct. at 2227. Under strict scrutiny, the highest standard of judicial review, the state can prevail only by showing an interest sufficiently compelling to make denying that right reasonable. Washington v. Glucksberg, 521 U.S. 702, 766–67 (1997) (Souter, J., concurring). In other words, the government must show that the infringement of that constitutional right or liberty interest is justified by a "compelling state interest" that the intrusion is "narrowly tailored to serve." Reno v. Flores, 507 U.S. 292, 301–02 (1993).

[74] *Snyder*, 562 U.S. at 453–54. For example, speech made in a public park or area, and thus consistent with traditional forms of public discourse, would receive greater protection than if it were made in a less traditional public forum or if legitimate government objectives justified its suppression. *See* Make the Rd. by Walking, Inc. v. Turner, 378 F.3d 133, 142 (2d Cir. 2004) (explaining that protection of speech is at its peak in a traditional public forum); Michael Kagan, *When Immigrants Speak: The Precarious Status of Non-Citizen Speech Under the First Amendment*, 57 B.C. L. REV. 1237, 1273 (2016) (discussing cases where the legitimate needs and objectives of government institutions require restricting speech). This is especially true for political speech, which receives heightened protection. *See Snyder*, 562 U.S. at 452–53 (discussing the heightened First Amendment protection for political speech due to the public nature of the issues); Citizens United v. FEC, 558 U.S. 310, 329 (2010) (declining to resolve a case in a

dresses issues of public concern rather than private concern, then it is generally afforded greater protection, as such speech is more valuable to the public discourse that the First Amendment seeks to promote and protect.[75] Speech is considered to address matters of public concern when it directly relates to any matter concerning the community, or when it involves a subject of "legitimate news interest," which is defined as a matter of common interest, significance, and concern to the public.[76] In evaluating whether speech deals with matters of public, rather than private, concern, the appropriateness or controversial character of the statement is not considered.[77]

The character of the speech, however, does play a role in categorizing the speech, which can also contribute to the level of First Amendment protection afforded.[78] The Supreme Court dictated categories of speech that offer so little social value that public interest favors depriving them of First Amendment protection.[79] Some categories, such as commercial speech, receive only limited First Amendment protection.[80] Others are fully excepted from its protection: fighting words, incitement to imminent lawless action, true threats, obscenity, child pornography, defamation, and speech integral to criminal conduct.[81] The first three fully excepted categories—fighting words, incitement to imminent lawless action, and true threats—are the focus of this Note, as they have all been invoked

---

manner that would chill political speech and emphasizing that the purpose of the First Amendment was largely to protect this speech).

[75] *See Snyder*, 562 U.S. at 452. Courts categorize matters that are only of personal interest to the speaker as private concerns. Connick v. Myers, 461 U.S. 138, 147 (1983). For example, a complaint by an employee insisting that he or she should have received a raise would "likely constitute a matter of only private concern and would therefore be unprotected" under the First Amendment. Janus v. AFSCME, 138 S. Ct. 2448, 2472–73 (2018).

[76] *Snyder*, 562 U.S. at 453 (quoting City of San Diego v. Roe, 543 U.S. 77, 83–84 (2004)).

[77] *Id.* (citing Rankin v. McPherson, 483 U.S. 378, 387 (1987)).

[78] *See* Planned Parenthood of Columbia v. Am. Coal. of Life Activists, 290 F.3d 1058, 1092 (9th Cir. 2002), as amended (July 10, 2002) (Reinhardt, J., dissenting) ("Speech—especially political speech, as this clearly was—may not be punished or enjoined unless it falls into one of the narrow categories of unprotected speech recognized by the Supreme Court: true threat, incitement, conspiracy to commit criminal acts, fighting words, etc.") (citations omitted).

[79] *R.A.V.*, 505 U.S. at 382–83 (quoting Chaplinsky v. New Hampshire, 315 U.S. 568, 572 (1942)); *see also* Terminiello v. City of Chicago, 337 U.S. 1, 4 (1949) ("[F]reedom of speech, though not absolute, is nevertheless protected against censorship or punishment, unless shown likely to produce a clear and present danger of a serious substantive evil that rises far above public inconvenience, annoyance, or unrest.").

[80] Hudson, *supra* note 66, § 2:5. Commercial speech is "expression related solely to the economic interests of the speaker and its audience." Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 561 (1980). This limited protection comes in the form of intermediate scrutiny standards of review, with suppression of such speech accepted only if it "directly advances" a "substantial" government interest in a manner that is only as extensive as necessary. *Id.* at 566.

[81] Hudson, *supra* note 66, § 2:5.

to either combat or protect hate speech.[82] Section B of this Part discusses these categories in greater detail.[83]

B. Categories of Unprotected Speech Applied to Hate Speech

In evaluating hate speech, courts often consider whether the speech falls into one of three related categories: incitement to imminent lawless action, fighting words, and true threats.[84] The first, incitement to imminent lawless action, emerged as a confluence of two tests the Court had previously applied to determine whether to afford the speech at issue First Amendment protection.[85] The latter two—fighting words and true threats—have been subject to relatively inconsistent application by courts, which have struggled to draw a clear line between protected speech and unprotected speech that fits within either category.[86]

1. Incitement to Imminent Lawless Action

Much of the First Amendment free-speech doctrine arose out of the government's attempts to silence dissidents that opposed the country's involvement in World War I.[87] One of the cases that arose during this period, *Schenck v. Unit-*

---

[82] *See, e.g.*, *Black*, 538 U.S. at 359–60 (evaluating cross-burning under the true threat framework after discussing true threats, fighting words, and incitement to imminent lawless action); D.C. v. R.R., 182 Cal. App. 4th 1190, 1200, 1219 (Cal. Ct. App. 2010) (analyzing students' hateful posts on a website directed towards and threatening another student under the true threat standard).

[83] *See infra* notes 84–124 and accompanying text.

[84] *See, e.g.*, *Black*, 538 U.S. at 359–60 (evaluating cross-burning under the true threat framework after discussing true threats, fighting words, and incitement to imminent lawless action); Brandenburg v. Ohio, 395 U.S. 444, 446 (1969) (evaluating statements made at Ku Klux Klan rally, including derogatory comments about blacks and Jews, under the incitement to imminent lawless action category); *Terminiello*, 337 U.S. at 3 (discussing speech containing critical remarks regarding certain political and racial groups and evaluating under the fighting words framework).

[85] Hudson, *supra* note 66, § 3:3.

[86] *Id.* §§ 3:10–:12. Federal courts have disagreed as to whether an objective or subjective standard should apply to true threats. *D.C. v. R.R.*, 182 Cal. App. 4th at 1213; *see, e.g.*, Fogel v. Collins, 531 F.3d 824, 831 (9th Cir. 2008) (explaining that the circuit previously applied a subjective standard of intent for true threats, but now considers both subjective and objective standards); United States v. Kosma, 951 F.2d 549, 557 (3d Cir. 1991) (favoring the objective, reasonable-person intent standard for threats, which is independent of the actual intent of speaker, and listing other circuits that have also taken this approach). Courts have also struggled to clearly ascertain the line between protected speech and fighting words. *Compare* Gower v. Vercler, 377 F.3d 661, 670 (7th Cir. 2004) (finding repeatedly telling neighbor "fuck you," calling neighbor a "fat-son-of-a bitch," and calling neighbor a coward to constitute fighting words), *with* Cornelious v. Brubaker, No. 01-1254, 2003 WL 21511125, at *2, *6 (D. Minn. June 25, 2003) (dismissing the idea that "fuck you" and similar profanities constitute fighting words).

[87] Hudson, *supra* note 66, § 3:2; *see also Am. Booksellers Ass'n*, 771 F.2d at 329 ("The Alien and Sedition Acts passed during the administration of John Adams rested on a sincerely held belief that disrespect for the government leads to social collapse and revolution—a belief with support in the history of many nations."). During this time, federal prosecutors used legislation such as the Espionage Act of 1917 and the Sedition Act of 1918 to silence political opposition to the war effort. Hudson, *supra* note 66, § 1:4; *see* Abrams v. United States, 250 U.S. 616, 619 (1919) (rejecting contention that First Amendment protects the speech that the defendants faced charges for under the Espionage Act of 1917). Various

ed States, involved the prosecution of two individuals for distributing documents criticizing the war effort and encouraging others to "[a]ssert [their] rights" and refuse to "submit to intimidation" with respect to the draft.[88] Justice Oliver Wendell Holmes, writing for the unanimous Supreme Court, emphasized the wartime context of the speech, which he found supported lesser protection.[89] In finding the speech unprotected, Justice Holmes introduced a new test—the "clear and present danger" test—which seemed to supplant the "bad tendency" test that courts had previously used to evaluate whether speech was criminal and, therefore, unprotected.[90]

Justice Holmes' new test asked whether the circumstances of the speech are of such a nature as to "create a clear and present danger that they will bring about the substantive evils that Congress has a right to prevent."[91] He further explained that the question is one of "proximity and degree" and demonstrated this by emphasizing that during times of war, certain speech that effectively hinders Congress's war effort may not be protected, even though it would have been afforded such protections during times of peace.[92] Holmes' test—which now, albeit modified by a later test, is accepted as the standard test for incitement speech—encompasses the principle that the government can only suppress unpopular political speech if it creates a clear, present danger to high-priority interests of the government, such as its interest in security.[93]

Following *Schenck*, the Supreme Court and Justice Holmes drew on Holmes' "clear and convincing danger" test to uphold convictions of Jacob Frohwerk and Eugene Debs, both of whom were charged after engaging in polit-

---

cases arose under these laws, and the Supreme Court reviewed a select few. Hudson, *supra* note 66, § 1:4; *e.g.*, *Abrams*, 250 U.S. 616; Debs v. United States, 249 U.S. 211 (1919); Frohwerk v. United States, 249 U.S. 204 (1919); Schenck v. United States, 249 U.S. 47 (1919).

[88] *Schenck*, 249 U.S. at 51; Hudson, *supra* note 66, § 1:4.

[89] *Schenck*, 249 U.S. at 52; Hudson, *supra* note 66, § 1:4.

[90] Hudson, *supra* note 66, § 1:4; *see Schenck*, 249 U.S. at 52. Under the prior "bad tendency" test, courts held that speech could be "penalized if it had a 'bad tendency' upon the public welfare." David M. Rabbant, *The First Amendment in Its Forgotten Years*, 90 YALE L.J. 514, 533 (1981); *see* Warren v. United States, 183 F. 718, 721 (8th Cir. 1910) (explaining that the ability of Congress to restrain the rights of liberty and freedom of speech, when doing so "in the interest of the general welfare, peace, and good order," is "beyond question" and consistently upheld by courts). This test was unfavorable to free speech, affording it little protection. *See* Hudson, *supra* note 66, § 3:3 ("This so-called 'bad tendency' test struck the balance heavily in favor of the government.").

[91] *Schenck*, 249 U.S. at 52; Hudson, *supra* note 66, § 3:3.

[92] *See Schenck*, 249 U.S. at 52 (elaborating on the circumstantial nature of the application of the test for clear and present danger).

[93] Hudson, *supra* note 66, § 1:4. Although by its terms "clear and present danger" may seem to have more of a stringent temporal requirement, the imminence requirement under this standard was not distinguishable from the "bad tendency" test previously used, indicating that this test only really added intent to the prior "bad tendency" test that allowed proscription of speech based on only vague or obscure proof of danger. 1 RODNEY A SMOLLA, SMOLLA AND NIMMER ON FREEDOM OF SPEECH § 10:4 (2018) (citing RODNEY A. SMOLLA, FREE SPEECH IN AN OPEN SOCIETY 99–100 (1992)).

ical speech that criticized the war effort.[94] Following these three speech-restrictive cases—*Schenck*, *Debs v. United States*, and *Frohwerk v. United States*—the Holmes Court pivoted, seeking instead to provide further protections for speech and a more rigorous construction of the standard for evaluating unprotected speech.[95] The Second World War and the Cold War, however, both accompanied an era of greater restriction on speech and lesser protection to political dissidents.[96] A pattern thus emerged, with greater freedom of speech during times of peace and greater restrictions on speech during times of war.[97]

In 1969, the standard was modified, adding definition to what Holmes referred to as "clear and convincing danger."[98] In *Brandenburg v. Ohio*, the Supreme Court introduced an incitement test, under which advocacy is denied constitutional protection where such speech "is directed to inciting or producing imminent lawless action and is likely to incite or produce such action."[99] This imminence requirement requires the government to show that the lawless action the speaker seeks to incite will result in immediate harm.[100] This so-called "*Brandenburg* incitement test" was later clarified in 1973, in *Hess v. Indiana*, where the Court applied the test to a protestor at Indiana University who threatened to "take the [] street."[101] The majority conceded that the protestor's speech,

---

[94] Hudson, *supra* note 66, § 3:3; *see Frohwerk*, 249 U.S. at 206 (rejecting First Amendment arguments and citing *Schenck* for support) (citing *Schenck*, 249 U.S. at 47); *Debs*, 249 U.S. at 215 (explaining that claims that Espionage Act of 1917 violated the First Amendment have been disposed of by *Schenck*) (citing *Schenck*, 249 U.S. at 47).

[95] Hudson, *supra* note 66, § 1:4; *see* Dennis v. United States, 341 U.S. 494, 505, 511 (1951) (applying a relatively broad version of clear and present danger test to uphold convictions of Communist party members during the Cold War); *Debs*, 249 U.S. 211 (speech-restrictive application of test); *Frohwerk*, 249 U.S. 204 (same); *Schenck*, 249 U.S. 47 (same).

[96] Hudson, *supra* note 66, § 3:3.

[97] *Id.*; *see also Dennis*, 341 U.S. at 528 (Frankfurter, J., concurring) (arguing that the constitutionality of convictions implicating First Amendment concerns "must be determined by principles established in cases decided in more tranquil periods" to avoid "the risk of an ad hoc judgment influenced by the impregnating atmosphere of the times"); Gilbert v. Minnesota, 254 U.S. 325, 338 (1920) ("There are times when those charged with the responsibility of Government, faced with clear and present danger, may conclude that suppression of divergent opinion is imperative; [sic] because the emergency does not permit reliance upon the lower conquest of error by truth. And in such emergencies the power to suppress exists.").

[98] *See Brandenburg*, 395 U.S. at 447 ("[T]he constitutional guarantees of free speech and free press do not permit a State to forbid or proscribe advocacy of the use of force or of law violation except where such advocacy is directed to inciting or producing imminent lawless action and is likely to incite or produce such action."); Hudson, *supra* note 66, § 3:2 (discussing the development of the *Brandenburg* test for incitement).

[99] *Brandenburg*, 395 U.S. at 447; Hudson, *supra* note 66, § 3:2 (quoting *Brandenburg*, 395 U.S. at 447).

[100] Hudson, *supra* note 66, § 3:2.

[101] Hess v. Indiana, 414 U.S. 105, 110 (1973); Hudson, *supra* note 66, § 3:5 (citing *Hess*, 414 U.S. 105); *see also* NAACP v. Claiborne Hardware, 458 U.S. 886, 928 (1982) ("An advocate must be free to stimulate his audience with spontaneous and emotional appeals for unity and action in a common cause. When such appeals do not incite lawless action, they must be regarded as protected speech.").

at best, involved advocating illegal conduct at an uncertain future time, but the Court could not find sufficient support—based on either extrinsic evidence or the words themselves—to justify finding that the speech was intended or likely to produce imminent disorder.[102] In the absence of a clear, imminent impact, the Court declined to deny the speech protection, emphasizing that a mere tendency to lead to violence was not enough to satisfy the imminence standard.[103] In doing so, the Court created imminence and likelihood requirements that distinguish this standard from prior ones.[104] Thus, modern jurisprudence tends to offer greater protection to speakers of inciting speech than prior jurisprudence, extending protection even to those advocating for lawless conduct or challenging authority, so long as such advocacy does not incite immediate lawless action.[105]

2. Fighting Words

Fighting words involve direct, personal insults that are likely to invoke a violent response out of the recipient.[106] Although similar to incitement speech, fighting words generally apply to negative remarks spoken directly to another individual, as opposed to remarks directed to a large group of people to incite them to engage in lawless activity that might harm one individual.[107] The Supreme Court first recognized fighting words in 1942, in *Chaplinsky v. New Hampshire*, where it upheld a New Hampshire statute as constitutional because it

---

[102] *Hess*, 414 U.S. at 108. The Court also emphasized that the speaker did not direct the speech towards any specific group or individual, which indicates that he was not advocating for action. *Id.* at 108–09.

[103] *Id.* at 109; Hudson, *supra* note 66, § 3:5. Under the *Brandenburg* test, courts analyze intent using an objective standard, which requires courts to consider objective facts and how one might reasonably interpret the speech in that context. United States v. White, 670 F.3d 498, 512 (4th Cir. 2012).

[104] Smolla, *supra* note 93, § 10:30; *see Am. Booksellers Ass'n*, 771 F.2d at 333 ("Cases . . . hold that a state may not penalize speech that does not cause immediate injury."). In applying the *Brandenburg* test, the speech itself and the circumstances surrounding the speech play a key role in distinguishing unprotected communications (e.g., communications relating to planning to set off a bomb in a public area at a definite time weeks in advance) and communications that fall short of the standard (e.g., communications that include rhetoric that tends to cause a crowd to lose control or endanger lives, but not specific plans of violent acts). Smolla, *supra* note 93, § 10:30.

[105] Hudson, *supra* note 66, § 3:5; *see* United States v. Fullmer, 584 F.3d 132, 155 (3d Cir. 2009) (finding website posts coordinating civil disobedience as unprotected speech and reasoning that urging individuals to participate in such disobedience at a set time "encouraged and compelled an imminent, unlawful act that was not only likely to occur, but provided the schedule by which the unlawful act was to occur"); *Am. Booksellers Ass'n*, 771 F.2d at 329 ("Most governments of the world act on this empirical regularity, suppressing critical speech. In the United States, however, the strength of the support for this belief is irrelevant. Seditious libel is protected speech unless the danger is not only grave but also imminent.").

[106] *Chaplinsky*, 315 U.S. at 573; Hudson, *supra* note 66, § 3:6.

[107] Hudson, *supra* note 66, § 3:6.

applied to "prohibit the face-to-face words plainly likely to cause a breach of the peace."[108]

The doctrine developed further in 1971, in *Cohen v. California*, where the Supreme Court rejected the argument that a man's jacket, which displayed the words "Fuck the Draft," involved fighting words, thus settling the proposition that fighting words must involve face-to-face insults, directly targeted at an individual.[109] The Supreme Court further explained the doctrine in *R.A.V. v. City of St. Paul*, where the Court emphasized that fighting words are denied First Amendment protection not because of their content, but rather because of the unwarranted and intolerable manner of expressing this speech.[110] Consequently, the government tends to take the position that a person charged with using fighting words was so charged based on conduct, such as yelling direct threats or flailing his or her arms, rather than the content of the speech or threat.[111]

3. True Threats

Like fighting words, true threats are another category of unprotected speech that has been the subject of varied interpretations by lower courts.[112] The Su-

---

[108] *Id.* (citing *Chaplinsky*, 315 U.S. at 573). Shortly after, in 1949, in *Terminiello*, the Supreme Court considered, but ultimately did not apply due to a preempting concern, this doctrine after Arthur Terminiello delivered a speech to a crowd of around eight hundred individuals. 337 U.S. at 3. In this speech, Terminiello "vigorously, if not viciously, criticized various racial and political groups," describing their activities as detrimental to national welfare. *Id.* Highlighting the fact that freedom of speech functions to invite dispute, the Supreme Court acknowledged that speech "may strike at prejudices and preconceptions and have profound unsettling effects as it presses for acceptance of an idea." *Id.* at 4. Because Justice Douglas did not think that the doctrine could apply to criminalize political speech, which the statute at issue did, he found further evaluation under the fighting words doctrine inappropriate. Hudson, *supra* note 66, § 3:8; *see Terminiello*, 337 U.S. at 3, 5 (declining to evaluate under the fighting words doctrine because a conviction based on the fact that "speech stirred people to anger, invited public dispute, or brought about a condition of unrest" cannot stand). Nevertheless, the Court explained, speech remains protected unless one can show that it is "likely to produce a clear and present danger of a serious substantive evil that rises far above public inconvenience, annoyance, or unrest." *Terminiello*, 337 U.S. at 4.

[109] Cohen v. California, 403 U.S. 15, 25 (1971); Hudson, *supra* note 66, § 3:8 (citing *Cohen*, 403 U.S. at 25).

[110] *R.A.V.*, 505 U.S. at 393.

[111] Hudson, *supra* note 66, § 3:10. *Compare* State v. Robinson, 82 P.3d 27, 28 (Mont. 2003) (yelling "fucking pig" unprovoked to a police officer constituted fighting words), *and Gower*, 377 F.3d at 670 (repeatedly calling a neighbor a "fat son of a bitch" and stating "fuck you" found to constitute fighting words), *with* Elbrader v. Blevins, 757 F. Supp. 1174, 1182 (D. Kan. 1991) (calling a police officer a "son of a bitch" did not constitute fighting words), *and Cornelious*, No. 01-1254, 2003 WL 21511125 (yelling "fuck you all" to a police officer deemed insufficient to constitute fighting words). For a related view, see Scott Hammack, Note, *The Internet Loophole: Why Threatening Speech On-line Requires a Modification of the Courts' Approach to True Threats and Incitement*, 36 COLUM. J.L. & SOC. PROBS. 65, 72–73 (2002) (explaining that, because some speech does not clearly fit into either the true threat or incitement standards, courts have misapplied these standards by interchangeably using them).

[112] *See* Hudson, *supra* note 66, §§ 3:10, 3:12 (discussing lower courts' applications of the standards for true threats and fighting words).

preme Court created this category in *Watts v. United States*, in 1969.[113] In *Watts*, the Court rejected claims that an eighteen-year-old's speech was a true threat, instead concluding that the speech was merely "a kind of very crude offensive method of stating a political opposition to the President."[114] In finding the speech protected, the Court emphasized the statement's conditional nature and the listeners' reactions, which apparently involved laughing at the statement.[115] The Court did not, however, define a "true threat," leading lower courts to focus instead on the so-called "*Watts* factors."[116] These factors include whether the statement accompanied a political debate; whether the threat was conditional; and whether the context of the speech, including listeners' reactions, is indicative of a true threat.[117]

After failing to provide a clear definition for true threat in another case, *NAACP v. Claiborne Hardware*, the Supreme Court finally provided a definition in *Virginia v. Black*, in 2003.[118] Justice Sandra Day O'Connor, in her plurality opinion, explained that true threats describe statements communicating an intent to commit some violent, unlawful act on a specified group or individual, regardless of whether an actual, subjective intent exists.[119] Thus, a definition of true threats, which depends upon the fear invoked and not the intent to carry out the threat, finally emerged.[120] Justice O'Connor further articulated that where the speaker directly targets an individual or group with the purpose of instilling fear of physical harm or death, the speaker engages in unlawful intimidation that amounts to a true threat.[121]

Even with this definition, confusion regarding what does and does not cross the line from protected speech to true threat remains.[122] Courts often apply an

---

[113] *Id.* § 3:12; *see* Watts v. United States, 394 U.S. 705, 707–08 (1969) (distinguishing a threat from constitutionally protected speech and referencing a "true 'threat'"). The speech at issue involved the young man's response to a comment made during a discussion on police brutality, in which the commenter suggested that the individuals present should seek further education before expressing their views on the topic. *Watts*, 394 U.S. at 706. In his response, Watts expressed his desire not to attend an upcoming physical relating to his recent draft and stated: "If they ever make me carry a rifle the first man I want to get in my sights is L.B.J. They are not going to make me kill my black brothers." *Id.*

[114] Hudson, *supra* note 66, § 3:12 (citing *Watts*, 394 U.S. at 707).

[115] *Id.* (citing *Watts*, 394 U.S. at 708).

[116] *Id.*

[117] *Id.*; *see* United States v. Mitchell, 812 F.2d 1250, 1255 (9th Cir. 1987) (citing *Watts*, 394 U.S. at 708) (listing these factors).

[118] Hudson, *supra* note 66, § 3:12; *see Black*, 538 U.S. at 359–60 (providing more definition); *Claiborne Hardware*, 458 U.S. 886.

[119] *Black*, 538 U.S. at 359–60; Hudson, *supra* note 66, § 3:12.

[120] *Black*, 538 U.S. at 359–60; Hudson, *supra* note 66, § 3:12. If the speaker only used the threatening language as political hyperbole or rhetoric this will not satisfy the "true threat" standard. Smolla, *supra* note 93, § 10:22.50 (citing *Watts*, 394 U.S. at 708).

[121] *Black*, 538 U.S. at 360.

[122] Hudson, *supra* note 66, § 3:12; *see also* Elonis v. United States, 135 S. Ct. 2001, 2028 (2015) (Thomas, J., dissenting) (criticizing majority's failure to resolve lack of clarity in true threat doctrine);

objective standard in evaluating whether speech constitutes a true threat, asking "whether a reasonable person would consider the statement a serious expression of an intent to inflict harm," although circuits have diverged in their analyses with regards to whether this perspective is one of a "reasonable speaker" or "reasonable listener."[123] In performing this analysis, courts consider the context of the threat as well, focusing on various factors, including the reactions of listeners, the nature of the threat (for example, conditional, direct, et cetera), any prior incidents where the speaker had threatened the victim, and any potential reasons for the recipient of the threat to believe that the speaker had violent propensities.[124]

Thus, while speech that may generally be classified as hate speech is not excepted from the First Amendment's protection, slivers of this speech may be denied this protection if they fall within the specific requirements of those already-established categories: fighting words, true threats, and incitement to imminent lawless action.[125] The result of this broadly-accepting policy on hate speech is an invigorated debate, requiring compromise of equally meritorious values, which is further explored in Part III.[126]

III. EXPLORING BOTH SIDES OF THE DEBATE

Critics of hate speech restrictions argue that regulating hate speech violates existing First Amendment principles, may lead to abuse through selective enforcement, and threatens to undermine the principles that underlie this country's liberal democracy.[127] Proponents of regulation disagree, emphasizing the harms

---

United States v. Syring, 522 F. Supp. 2d 125, 129 (D.D.C. 2007) (providing examples to demonstrate this split).

[123] *Syring*, 522 F. Supp. 2d at 129. In 2015, the Supreme Court had the opportunity to resolve the intent standard of true threats, but failed to do so. *See Elonis*, 135 S. Ct. at 2028 (Thomas, J., dissenting) ("Given the majority's ostensible concern for protecting innocent actors, one would have expected it to announce a clear rule—any clear rule. Its failure to do so reveals the fractured foundation upon which today's decision rests."); *see also* Stephanie Charlin, Comment, *Clicking the "Like" Button for Recklessness: How* Elonis v. United States *Changed True Threats Analysis*, 49 LOY. L.A. L. REV. 705, 707 (2016) (discussing the growing importance of resolving intent issue in the digital age, where threats can more easily and rapidly inflict harm than ever before, and criticizing the Court's failure to resolve the issue in *Elonis*).

[124] *Syring*, 522 F. Supp. 2d at 130 (citing United States v. Dinwiddie, 76 F.3d 913, 925 (8th Cir. 1996)).

[125] *See Black*, 538 U.S. at 359 (discussing the exclusion of fighting words, incitement to imminent lawless action, and true threats); *see, e.g.*, D.C. v. R.R., 182 Cal. App. 4th at 1200, 1225 (analyzing students' hateful posts on a website directed towards and threatening another student under true threat standard).

[126] *See infra* notes 127–202 and accompanying text.

[127] *See, e.g.*, Sandra Coliver, *Hate Speech Laws: Do They Work?*, *in* STRIKING A BALANCE: HATE SPEECH, FREE SPEECH, AND NON-DISCRIMINATION 263, 374 (Sandra Coliver ed., 1992) (describing the harmful effects of "selective or lax enforcement" in various nations and questioning the effectiveness of hate speech regulations in Europe based on increases in racist and xenophobic sentiments in European

advanced by hate speech, referencing the regulations imposed in other democracies, and noting the outdated origins of First Amendment jurisprudence.[128] Section A of this Part explores the arguments advanced by regulation critics in more detail.[129] Section B explores those maintained by advocates of hate speech regulation.[130]

A. Arguments Against Hate Speech Regulation

Critics of hate speech regulation typically make policy arguments warning against violation of First Amendment principles and practical arguments based on trends in countries that regulate such speech, focusing on reports of abuse by those charged with enforcing speech restrictions and on the failure of regulations to effectively combat hate.[131]

1. First Amendment Arguments

First, critics argue that speech of this nature should be evaluated as incitement to imminent lawless action, which requires a showing of clear and present danger that generally cannot be made with hate speech.[132] In doing so, critics

countries); *see also* Nadine Strossen, *Incitement to Hatred: Should There Be a Limit?*, 25 S. ILL. U. L.J. 243, 259 (2001) (citing HUMAN RIGHTS WATCH, *'Hate Speech' and Freedom of Expression, A Human Rights Watch Policy Paper*, Mar. 1992, at 4) ("The conclusion of all these papers was clear: not even any correlation, let alone any causal relationship, could be shown between the enforcement of anti-hate speech laws by the governments in particular countries and an improvement in equality or inter-group relations in those countries.").

[128] *See, e.g.*, Calvin R. Massey, *Hate Speech, Cultural Diversity, and the Foundational Paradigms of Free Expression*, 40 UCLA L. REV. 103, 155 (1992) (explaining that the potential to incite violence and harm society generally is a common justification for suppression of hate speech); Tsesis, *supra* note 14, at 861 ("Germany is another democracy committed to free expression which, nevertheless, recognizes the social menace posed by hate speech and penalizes it."); Hammack, *supra* note 111, at 81 (emphasizing that the relevant case law developed prior to the rise of Internet communications and listing the characteristics that differentiate harmful Internet communications from those made in a more traditional forum).

[129] *See infra* notes 131–147 and accompanying text.

[130] *See infra* notes 148–205 and accompanying text.

[131] *See generally* John T. Bennett, *The Harm in Hate Speech: A Critique of the Empirical and Legal Bases of Hate Speech Regulation*, 43 HASTINGS CONST. L.Q. 445 (2016) (arguing against the regulation of hate speech by presenting policy arguments and practical arguments based on uncertainties about intangible, speech-based harms).

[132] *See* Strossen, *supra* note 127, at 244–45 ("Our position is not that government may never restrict speech, but rather, that it may do so only under very limited circumstances. In a nutshell, government may suppress speech only if necessary to prevent a clear and present danger of actual or imminent harm."); *accord* Bennett, *supra* note 131, at 476, 500 ("Speech regulation requires more than a loose connection or imaginary association between speech and social harm. Whether hate speech causes imminent harm is an empirical question with major constitutional ramifications."); *see also* Schenck v. United States, 249 U.S. 47, 52 (1919) (standard for incitement speech). Scholars have also considered hate speech in the context of fighting words and true threats, but the nature of hate speech makes analysis most appropriate under the incitement standard. *See* Bennett, *supra* note 131, at 525 ("A theoretical 'climate of hate' will not justify restrictions on speech because hate falls short of harm. In fact, hate falls short of being a threat as well."); William Funk, *Intimidation and the Internet*, 110 PENN ST. L. REV.

reject arguments involving the harm of hate speech, either as too attenuated from the speech itself or insufficiently severe to justify a finding of a clear and present danger of imminent harm based on that connection alone.[133] Some have even challenged the notion that banning hate speech would be beneficial, suggesting that shielding individuals from speech that is associated with negative psychological impacts can actually undermine mental health and may even lead to retaliatory violence by the speakers.[134] Accordingly, with this showing hindered by the apparently weak causal connection between hate speech and social harm, critics argue that one cannot properly fit hate speech within a categorical exclusion from the First Amendment.[135]

Second, and relatedly, scholars insist that banning hate speech where there is no clear and present danger of imminent harm amounts to unconstitutional viewpoint discrimination, as the speech does not fall into an excluded category of speech.[136] Critics rely on the "marketplace of ideas" theory to defend their commitment to the First Amendment and its protection of hate speech, explaining that a free market of ideas is necessary for the truth to prevail.[137] Relatedly, they insist that a plurality of views must be available for the promotion of democracy through public discourse, as this allows individuals to access opposing

---

579, 580 (2006) (discussing the limitations of existing categories due to the nature of hate speech). As Scott Hammack explains, hate speech on the Internet often takes the form of "threat/incitement hybrids" that maintain First Amendment protection because they can neither clearly be considered incitement nor true threats. Hammack, *supra* note 111, at 67. These hybrids, he explains, are a result of the unique characteristics of the Internet converging to blur the line between threats and incitements by allowing people to threaten through this incitement. *Id.* In other words, speech on the Internet allows individuals to create fear by increasing the risk of ensuing violence, without actually making any explicit threats themselves. *Id.*

[133] *See, e.g.*, Bennett, *supra* note 131, at 476, 500 (attacking harm-based arguments based on a weak causal connection and based on insufficient harms); Strossen, *supra* note 127, at 250 (attacking harm-based arguments based on a weak causal connection).

[134] *See, e.g.*, NADINE STROSSEN, HATE: WHY WE SHOULD RESIST IT WITH FREE SPEECH, NOT CENSORSHIP 150–51 (Geoffrey R. Stone et al. eds., 2018) (warning that eliminating exposure to hate speech might harm mental health by reducing beneficial psychological stress and suggesting that government suppression of speech may result in retaliatory violence by speakers); *accord* Bennett, *supra* note 131, at 500 ("Rather than advancing social justice, calls for regulation may actually entrench the status quo, which uplifts no one and is actually debilitating.").

[135] Bennett, *supra* note 131, at 525.

[136] *See, e.g.*, Strossen, *supra* note 127, at 252 (describing content or viewpoint neutrality as a fundamental principle of free speech that suppression of hate speech violates). Thus, critics argue that regulation would violate the principle of viewpoint neutrality, which requires that the state refrain from restricting public speech based on disagreement with the view expressed, unless the suppression of speech is narrowly tailored to serve a compelling state interest. *See id.* (discussing viewpoint neutrality, the appropriate standard of scrutiny applied to viewpoint discrimination, and the risks that this principle mitigates); *accord* Bennett, *supra* note 131, at 505–07 (evaluating hate speech restrictions under the standard applied to content-based speech restrictions and concluding that such regulations inevitably will fail to meet the narrow tailoring requirement).

[137] Victor C. Romero, *Restricting Hate Speech Against "Private Figures": Lessons in Power-Based Censorship from Defamation Law*, 33 COLUM. HUM. RTS. L. REV. 1, 16 (2001); *see, e.g.*, Weintraub-Reiter, *supra* note 3, at 162 (justifying and explaining the marketplace of ideas theory).

views and determine which one they deem correct.[138] Emphasizing the value of these opposing views, or "counter-speech," critics warn about the dangers of interfering with these democratic processes by weakening protection or stretching the doctrine in an attempt to find hate speech unprotected.[139]

2. Practicality & Practice Arguments

Arguments grounded in practice, rather than policy, are also common among critics of regulation.[140] For example, critics of hate speech regulation emphasize potentials for abuse that arise with such regulation.[141] Critics warn of selective enforcement and "flagrant abuse" by authorities in various countries, arguing that hate speech restrictions in these areas have been used to the detriment of minority communities, leading to alienation and compromise of the right of dissent.[142] Some have drawn on this to urge hate speech restriction advocates to focus instead on promoting non-discrimination more generally, reasoning that

---

[138] *See, e.g.*, Weintraub-Reiter, *supra* note 3, at 161–62 (insisting that allowing hate speech promotes democracy by providing a plurality of views from which one may assess and utilize to engage in informed democratic decision making).

[139] *See* Bennett, *supra* note 131, at 502 (discussing the dangers of stretching First Amendment doctrine too far and insisting that its foundation will progressively weaken as more and more speech loses protection) (quoting Mari J. Matsuda, *Public Response to Racist Speech: Considering the Victim's Story*, 87 MICH. L. REV. 2320, 2357 (1989)); *accord* Weintraub-Reiter, *supra* note 3, at 162 ("[T]he restriction of speech on media would ultimately inhibit the growth of a democratic society."); Erwin Chemerinsky, *Commentary: Hate Speech Is Infecting America, but Trying to Ban It Is Not the Answer*, CHI. TRIB. (Oct. 31, 2018), https://www.chicagotribune.com/news/opinion/commentary/ct-perspec-hate-speech-censor-first-amendment-1101-20181031-story.html [https://perma.cc/D2PQ-EZQY] (contending that granting governments the power to suppress speech they dislike would be more harmful than hate speech).

[140] *See, e.g.*, STROSSEN, *supra* note 134, at 81 ("Given the pervasiveness of individual and institutional bias, the government is likely to enforce 'hate speech' laws, as it has other laws, to the disadvantage of disempowered and marginalized groups. Indeed, laws censoring 'hate speech' have predictably been enforced against those who lack political power . . . ."); *see also* Coliver, *supra* note 127, at 373–74 (discussing abuse of laws restricting hate speech by authorities in Sri Lanka, South Africa, Eastern Europe, and the former Soviet Union and selective enforcement in United Kingdom, Israel, and the former Soviet Union); Strossen, *supra* note 127, at 258–59 (discussing findings that indicate the ineffectiveness of hate speech restrictions in combating social inequality, bias, and discrimination).

[141] *See, e.g.*, Strossen, *supra* note 127, at 258 ("Laws that penalize speech or membership are also subject to abuse by the dominant racial or ethnic group. Some of the most stringent 'hate speech' laws, for example, have long been in force in South Africa, where they have been used almost exclusively against the black majority.") (citing HUMAN RIGHTS WATCH, *supra* note 127, at 4). Many attribute this potential for abuse to the subjective nature of delineating what speech amounts to hate speech, which is heightened by the ambiguity in defining hate speech. *See* STROSSEN, *supra* note 134, at 72 (explaining that the subjective nature of inquiry behind discerning hate speech increases its potential for abuse); Bennett, *supra* note 131, at 487 (suggesting that relying on psycho-emotional harms to evaluate whether speech should be curtailed "would call for a highly subjective inquiry into personal feelings" and noting the constitutional concerns it would raise).

[142] *See* Coliver, *supra* note 127, at 373–74 (warning of abuse); Strossen, *supra* note 127, at 259 (discussing Coliver's findings and concerns). Sandra Coliver, for example, points to selective enforcement in the United Kingdom, Israel, and the former Soviet Union, and she notes abuse by authorities in Sri Lanka and South America. Coliver, *supra* note 127, at 374; Strossen, *supra* note 127, at 259.

racism and xenophobia have become increasingly prevalent throughout Europe despite the existence of such hate speech laws.[143]

Similar arguments against speech restriction emphasize the failure of such regulations to combat discrimination and intolerance after implementation.[144] Critics rely on various studies to demonstrate this ineffectiveness, such as the findings of international free speech organization Article 19 after hosting a conference in 1991, during which representatives from fifteen countries gathered to discuss and evaluate the effectiveness of their respective anti-hate speech laws in combating discrimination, bias, and inequality.[145] The findings failed to show any causal or correlative relationship between enforcement and improvements in combatting social inequality and discrimination.[146] Critics have also pointed to another experience and observation-based study, conducted in 1992, in which international human rights organization Human Rights Watch concluded that suppressing hate speech is an ineffective means of promoting equality and lessening discrimination, a conclusion based on its finding of a weak connection between suppression and reduced ethnic or racial tension.[147]

B. Hate Speech Regulation Advocates' Arguments

While recognizing the First Amendment principles at stake, proponents of hate speech regulation insist that a complete compromise of democratic institutions is avoidable and, to the degree that any compromise must be made, it is warranted.[148] Advocates point to other democracies that have also sought to balance the right to free speech with the desire to preserve democratic institutions

---

[143] *See* Coliver, *supra* note 127, at 373–74 (questioning the effectiveness of regulations and pointing to increases in bias and discrimination despite European hate speech laws); *accord* STROSSEN, *supra* note 134, at 137 (discussing a survey among European Jews, wherein sixty-seven percent reported increases in anti-Semitism despite the enactment of hate speech laws).

[144] *See, e.g.*, Strossen, *supra* note 127, at 258–59 (discussing findings of Article 19 and Human Rights Watch).

[145] *See, e.g.*, *id.* (relying on this study for support). Article 19 seeks to promote freedom of expression and strives for a world in which speakers may enjoy this protection in the absence of fear of discrimination. *Annual Report*, ARTICLE 19, https://www.article19.org/about-us/annual-report/ [https://perma.cc/5YNW-93XF]. In doing so, its team works on local, national, and international levels to "promote media freedom, increase access to information, protect journalists and human rights defenders, fight the shrinking of civic space, and place human rights at the heart of developing digital spaces." *Id.*

[146] Strossen, *supra* note 127, at 259.

[147] *See id.* at 258 (discussing the Human Rights Watch's study and findings).

[148] *See* Matsuda, *supra* note 139, at 2630–31 ("While the value of free speech can guide the choice of procedure—including evidentiary rules and burdens of persuasion—it should not completely remove recourse to the institution of law to combat racist speech."); Tsesis, *supra* note 14, at 869 ("[A]bstract uncertainties about potential evils should not constrain legislators from passing laws narrowly designed to curb expressions whose only object is to endanger the lives, professions, properties, and civil liberties of the less powerful."); *see also* Tsesis, *supra* note 14, at 858 (explaining that other countries have enacted legislation to suppress hate speech in recognition of the threats hate speech poses to democracies, human rights, and human dignity).

but have nevertheless adopted laws against hate speech, such as Germany and Canada.[149] In making their own arguments in favor of regulation, advocates focus on the antiquity of the current First Amendment doctrine—the foundation of which lies upon assumptions that they contend translate poorly into the Internet age—and on the harms created by both online and offline hate speech.[150]

1. Attacks on an Antiquated Doctrine

Advocates of regulation have criticized existing First Amendment jurisprudence as outdated, highlighting the antiquated presumptions built within its foundation.[151] Much of this criticism is due in large part to the drastic changes brought on by the Internet and amplified by social media, both of which possess certain characteristics that complicate application of First Amendment jurisprudence.[152] To demonstrate this complication, advocates of regulation emphasize the unique qualities of cyberspace.[153] For example, cyberspace allows for both

---

[149] *See* Timofeeva, *supra* note 25, at 254, 257 (contrasting the United States' approach with that of Germany and explaining that both nations share commitment to freedom of speech and traditional liberalism tenets); Tsesis, *supra* note 14, at 861–62 (discussing Canada, Germany, and other western democracies that have adopted hate speech restrictions); Thomas J. Webb, Note, *Verbal Poison—Criminalizing Hate Speech: A Comparative Analysis and a Proposal for the American System*, 50 WASHBURN L.J. 445, 446 (2011) (contrasting the United States' position on hate speech with those of "nearly every nation across the globe" who have regulated hate speech in favor of promotion of human dignity and protection of minorities over freedom of speech).

[150] *See, e.g.*, Bennett, *supra* note 131, at 475 (explaining the harm argument made by advocates of hate speech regulation); Hammack, *supra* note 111, at 81 (explaining cases behind the development of speech that incites imminent lawless action and noting that all of the cases previously discussed "predate the Internet's emergence as a popular mode of communication").

[151] *See* Funk, *supra* note 132, at 580 (arguing that First Amendment doctrine does not contemplate, and thus does not directly address, the type of threatening and inciting speech communicated on the Internet); Lyrissa Barnett Lidsky, *Incendiary Speech and Social Media*, 44 TEX. TECH L. REV. 147, 161 (2011) (discussing the problem with applying *Brandenburg* standard to social media and highlighting differences between audiences and speakers contemplated by the *Brandenburg* test and those who use social media to express views); Hammack, *supra* note 111, at 66 (explaining that First Amendment jurisprudence relating to potentially threatening speech arose in the context of communications made in "fundamentally different media," where cases typically involved "remarks relayed to a very limited audience through pamphlets or at small rallies" and thus required a simpler analysis).

[152] *See* Lidsky, *supra* note 151, at 161 (criticizing First Amendment doctrine for this reason); Hammack, *supra* note 111, at 96 (explaining that traditional approaches to true threats fail to effectively combat online threats and noting that the Internet aggravates pre-existing doctrinal shortcomings and problems); *see also* Timofeeva, *supra* note 25, at 254 ("In spite of many new communicative and technical options of the Internet, both the United States and Germany attempt to fit this new media into their old free speech standards.").

[153] *See* Timofeeva, *supra* note 25, at 253–54 (describing unique qualities of the Internet, including its various communicative options in terms of parties involved and audience number, its lack of inherent restrictions on size or resources, and its provision of globalism and anonymity); Thomas E. Crocco, Comment, *Inciting Terrorism on the Internet: An Application of* Brandenburg *to Terrorist Websites*, 23 ST. LOUIS U. PUB. L. REV. 451, 456 (2004) ("[T]he Court has not yet fully explored the unique characteristics of a ubiquitous electronic forum, where speakers are unseen and listeners unknown in a non-

contextual, geographical, and temporal dislocations, as anything posted on the Internet can have an international reach and can be viewed long after it is originally posted.[154] Especially with the advent of social media, cyberspace has drastically increased the prospective audience size and the number of individuals able to participate in unmediated, unregulated communication.[155] Both the audience and speaker can remain anonymous, and the physical crowd that typically comprises an audience is rare.[156] Advocates also note that, whereas traditional forms of communications forced one to pause and reflect before response, the Internet permits much more immediate results that can lead to more violent or instinctive reactions.[157]

In contrast, much of First Amendment doctrine developed in an age where physical space, geographical location, and time limited an audiences' size and composition.[158] Advocates highlight consequences stemming from these differences, arguing, for example, that counter-speech, a common justification for the United States' protection of most speech, is unlikely to exist within the like-minded communities that form through online communications, making counter-speech an ineffective deterrent.[159] Additionally, the inability to ascertain the pre-

---

contemporaneous setting."); Hammack, *supra* note 111, at 67 ("The unique characteristics of the Internet blur the distinction between threats and incitement by allowing speakers to threaten by incitement . . . .").

[154] Lidsky, *supra* note 151, at 148–49; *see also* Hammack, *supra* note 111, at 67 ("[T]he Internet allows a potentially unlimited and transient audience to communicate across the world with great speed and anonymity, and to do so at a fraction of the cost of other modes of communication.").

[155] Lidsky, *supra* note 151, at 149.

[156] *Id.*; *see also id.* ("A social media audience member is truly part of a lonely crowd.") (quoting Janet Morahan-Martin & Phyllis Schumacher, *Loneliness and Social Uses of the Internet*, 19 COMPUTERS HUM. BEHAV. 659, 660 (2003)).

[157] *See, e.g.*, Hammack, *supra* note 111, at 83 (discussing traditional communication's longer period of time for deliberation and self-restraint due to delay between generating a thought and subsequently sharing that thought, which the Internet has reduced through its facilitation of low-cost, high-speed communication). As Lyrissa Barnett Lidsky explains, the audience or "crowds" on social media do not maintain "the physical connections between crowds in 'real space' [that] potentially exert a restraining influence on the individual who is spurred to violent actions by the words of a fiery speaker." Lidsky, *supra* note 151, at 149–50.

[158] Lidsky, *supra* note 151, at 149. Prior to the advent of the Internet, leading cases on threatening speech involved remarks made to limited audiences either through pamphlets or small rallies. Hammack, *supra* note 111, at 66. As Thomas E. Crocco explains, cyberspace has "replaced the soapbox as the 'poor man's' forum," thus replacing the contemporaneous settings contemplated by the doctrine with non-contemporaneous ones "where speakers are unseen and listeners unknown." Crocco, *supra* note 153, at 456.

[159] *See, e.g.*, Hammack, *supra* note 111, at 82 (explaining that "the inability to ensure that the public has access to both sides of the debate" on the Internet often makes a productive debate involving counter-speech impossible). Hammack provides an example to illustrate this concept, explaining that "if an anti-Semitic web site publishes falsehoods slandering Jews, visitors to that site would be unlikely to visit a Jewish organization's web site refuting the anti-Semitic speech." *Id.* He emphasizes the widely scattered nature of an online audience and contrasts that with those of more traditional forms of media, such as pamphlets, televisions, and publications. *See id.* at 81–82.

cise audience of an Internet communication at any given time hinders the ability to fight any speech with counter-speech or rational debate.[160]

According to regulation advocates, attempting to fit online hate speech into the doctrines designed to combat similarly socially-undesirable, threatening, or harm-inciting speech demonstrates First Amendment doctrine's inability to contemplate those changes.[161] For example, because fighting words contemplate direct, face-to-face interaction, that doctrine is unable to accommodate direct speech that is intended to serve the same purpose, but communicated online, *i.e.*, face-to-screen-to-screen-to-face communication.[162] The true threat doctrine also falls short in the context of Internet hate speech because hate speech takes the form of more indirect threats, either implicit through the threats of harm it poses to its intended audience or victims, or explicit but directed to a general group of people rather than a specific target.[163]

Attempts to fit this speech into the incitement to imminent lawless action category also fail, a consequence of the imminence requirement.[164] The temporal, geographic, and contextual dislocation made possible by cyberspace makes any direct, immediate consequence of an Internet communication—whether harmful or not—almost impossible to ascertain.[165] It is difficult to iden-

---

[160] *Id.* at 81–82.

[161] *See, e.g.*, Crocco, *supra* note 153, at 457 ("*Brandenburg* provides the basis for making that determination for speech that incites others to unlawful activity, but its modern application has been to situations more akin to the real-time characteristics of a soapbox than to the virtual, extra-contemporaneous character of the Internet."); Hammack, *supra* note 111, at 67 (discussing the difficulty in applying true threat doctrine to online hate speech and arguing that courts should refine their approach to true threats to adapt to the changes brought about by the Internet); *see also* John P. Cronan, *The Next Challenge for the First Amendment: The Framework for an Internet Incitement Standard*, 51 CATH. U. L. REV. 425, 456 (2002) (arguing that the "goals of preventing the undesirable consequences of incitement" cannot be attained on the Internet without altering the interpretation of the imminence requirement).

[162] *See* Hudson, *supra* note 66, § 3:6 (explaining the direct, face-to-face nature of the type of speech that gave rise to the fighting words doctrine); *see also* R.A.V. v. City of St. Paul, 505 U.S. 377, 402 (1992) (White, J., concurring) (criticizing the majority for legitimizing hate speech as a form of debate by imposing a standard for fighting words contingent on conduct and context, rather than content of those fighting words alone). Justice White expresses disagreement with Supreme Court jurisprudence on fighting words for this exact reason, arguing that "a ban on all fighting words or on a subset of the fighting words category would restrict only the social evil of hate speech, without creating the danger of driving viewpoints from the marketplace." *R.A.V.*, 505 U.S. at 401.

[163] Hammack, *supra* note 111, at 67.

[164] Crocco, *supra* note 153, at 456. For example, a few weeks before the election of President Barack Obama, Walter Bagdasarian posted the following statements online: "'Re: Obama fk the niggar, he will have a 50 cal in the head soon' and . . . 'shoot the nig.'" United States v. Bagdasarian, 652 F.3d 1113, 1115 (9th Cir. 2011). Judge Stephen Roy Reinhardt, speaking for the majority, described the statements as "particularly repugnant because they directly encourage violence," but found the statements constitutionally protected "because urging others to commit violent acts 'at some indefinite future time' does not satisfy the imminence requirement for incitement under the First Amendment." *Id.* at 1115 n.9 (citing Hess v. Indiana, 414 U.S. 105, 108 (1973)).

[165] Lidsky, *supra* note 151, at 148–49. This contributes to issues in enforcement of hate speech bans in other countries, which the Ninth Circuit had to face when asked to uphold a French court's order, pursuant to the French Criminal Code, that required Yahoo! to remove French citizens' access to Nazi

tify any immediate responses because it is challenging to identify a precise audience.[166] This difficulty is due both to the size of a potential audience and the fact that a different audience may view the same post at different times, from different perspectives, in different contexts, and in different parts of the world.[167] Thus, a post can directly cause incitement, but the imminence of that incitement is obscured by those dislocations.[168] Regulation advocates, recognizing these shortcomings, urge courts to take action so that the doctrine can successfully combat harmful speech in such scenarios, emphasizing the need for flexibility in a doctrine meant to adapt to various scenarios and contexts.[169]

2. Harm and Dignity Arguments

Advocates argue that the harms stemming from both online and offline hate speech should warrant protection, despite the First Amendment implications of such protection.[170] They point to the harmful physical, mental, and social impacts that messages transmitted through the Internet have on targeted individuals and groups.[171] Drawing on studies finding these harms associated with hate

---

propaganda presented on its website. *See* Timofeeva, *supra* note 25, at 275. Yahoo!, a California-based company, argued that this would violate the First Amendment. Yahoo! Inc. v. La Ligue Contre Le Racisme, 433 F.3d 1199, 1206 (9th Cir.) (en banc) (per curiam), *cert. denied*, 547 U.S. 1163 (2006). After three judges concluded that the suit was "unripe for decision" and three dissenting judges concluded that the district court did not have personal jurisdiction, the Ninth Circuit remanded to the district court to dismiss the suit without prejudice. *La Ligue Contre Le Racisme*, 433 F.3d at 1224. See Timofeeva, *supra* note 25, at 275, for a discussion of this case.

[166] Hammack, *supra* note 111, at 81.

[167] Lidsky, *supra* note 151, at 148–49.

[168] *See id.* (discussing the difficulties of applying incitement standard to online posts without subverting doctrine's goals).

[169] *See, e.g.*, Charlin, *supra* note 123, at 707 ("Given the speed and ease with which online threats can inflict harm, this long-disputed question of intent needs a quick resolution, as demonstrated by the number of cases involving true threats now percolating in the courts."); Crocco, *supra* note 153, at 456 (discussing the need for flexibility in this doctrine).

[170] *See* Massey, *supra* note 128, at 155 (asserting that protection of individual interests against invasion and preservation of "a more general societal interest in preventing violent rupture of social norms" justifies denying constitutional protection); Matsuda, *supra* note 139, at 2360 ("What is argued here . . . is that we accept certain principles as the shared historical legacy of the world community. Racial supremacy is one of the ideas we have collectively and internationally considered and rejected . . . . We are not safe when these violent words are among us.").

[171] *See* Matsuda, *supra* note 139, at 2332 ("In addition to physical violence, there is the violence of the word. Racist hate messages, threats, slurs, epithets, and disparagement all hit the gut of those in the target group."); *id.* at 2377 ("As Professor Delgado has noted, the underlying first amendment values of self-fulfillment, knowledge, participation, and stable community recognized by first amendment theorists are sacrificed when hate speech is protected.") (citing Richard Delgado, *Words That Wound: A Tort Action for Racial Insults, Epithets, and Name-Calling*, 17 HARV. C.R.-C.L. L. REV. 133 (1982)); Tsesis, *supra* note 14, at 863–64 ("What is needed is a legal scheme to regulate the Internet because the messages transmitted through that social space have physical, psychological, and cultural effects on real places and real people."). *See generally* Delgado, *supra* (discussing psychological, physical, and societal harms caused by permitting hate speech).

speech, proponents of regulation insist that regulation can combat and reduce those harms.[172] Moreover, regulation advocates argue that the harm perpetuated by hate speech extends further than that suffered by the targeted individuals; it extends to society more generally.[173]

Advocates of hate speech regulations have identified two harms in particular: one direct and one indirect, both of which become clear in the context of hate speech on a white supremacist website.[174] The direct harm of hate speech is that it contributes to humiliation and degradation suffered by targeted groups, particularly minorities.[175] Advocates explain that racial insults, a common form of hate speech, are intentional affronts to personal dignity that lead to both immediate emotional distress and long-term emotional pain that contributes to the psychological harm caused by stigmatization and disrespect suffered by those victims.[176] Drawing on the findings of social scientists on the subject, advocates emphasize that degrading stereotypes can become "self-fulfilling prophecies" because racial insults are constantly directed towards these individuals.[177] Such

---

[172] *See, e.g.*, Delgado, *supra* note 171, at 146 (discussing social scientists' studies on the effects of racism) (citing M. DEUTSCH ET AL., SOCIAL CLASS, RACE AND PSYCHOLOGICAL DEVELOPMENT 175 (1968)); Matsuda, *supra* note 139, at 2361 ("Racism as an acquired set of behaviors can be dis-acquired, and law is the means by which the state typically provides incentives for changes in behavior."); Tsesis, *supra* note 14, at 869 (citing GORDON W. ALLPORT, THE NATURE OF PREJUDICE 57 (25th Anniversary ed. 1979)) (explaining Allport's theory that, because discriminatory laws increase prejudice, laws prohibiting discrimination should increase tolerance and describing him as "one of the foremost experts on the psychology of bigotry"); *see also* Bennett, *supra* note 131, at 447 ("With a remarkable degree of uniformity, calls for hate speech regulation rest on supposed social harms or inequalities, and presume that severe and widespread speech-based harm is a frequent aspect of life with a constitutionally significant impact on minorities.").

[173] *See* Delgado, *supra* note 171, at 140 ("Racism and racial stigmatization harm not only the victim and the perpetrator of individual racist acts but also society as a whole. Racism is a breach of the ideal of egalitarianism, . . . an ideal that is a cornerstone of the American moral and legal system."); N. Douglas Wells, *Whose Community? Whose Rights?—Response to Professor Fiss*, 24 CAP. U. L. REV. 319, 319 (1995) ("The harm caused by hate speech is greater than the psychological harm to the victims of hate speech; it also includes harm to society at large.").

[174] *See* David Kretzmer, *Freedom of Speech and Racism*, 8 CARDOZO L. REV. 445, 462 (1987) ("Two arguments for restricting racist speech are based on different visions of the types of harm it engenders: spread of racial prejudice, and affront to personal dignity."); Romero, *supra* note 137, at 6, 8 (discussing the direct harms and indirect harms recognized by advocates).

[175] Delgado, *supra* note 171, at 146; Romero, *supra* note 137, at 6; *see also* ALLPORT, *supra* note 172, at 142 ("One's reputation, whether false or true, cannot be hammered . . . into one's head without doing something to one's character.").

[176] *See* Delgado, *supra* note 171, at 143, 145–46 (discussing dignitary, emotional, and other psychological harms associated with hate speech and emphasizing the malicious intent behind hate speech); Tsesis, *supra* note 14, at 842 (discussing dignitary harms); *see also* Charles R. Lawrence III, *Crossburning and the Sound of Silence: Antisubordination Theory and the First Amendment*, 37 VILL. L. REV. 787, 796–97 (1992) ("The primary intent of the cross burner in *R.A.V.* was not to enter into a dialogue with the Jonses, or even with the larger community . . . . His purpose was to intimidate . . . . The discriminatory impact of this speech is of even more importance than the speaker's intent.").

[177] *See* Delgado, *supra* note 171, at 146 (quoting M. DEUTSCH ET AL, *supra* note 172, at 175) (discussing prophetic nature of racial insults); *accord* Am. Booksellers Ass'n, Inc. v. Hudnut, 771 F.2d 323,

insults either implicitly or explicitly communicate those stereotypes.[178] Psychological responses to this phenomenon, including self-hatred, humiliation, and isolation, lead those stigmatized individuals "to feel ambivalent about their self-worth and identity."[179] Advocates urge society to recognize this dignitary harm, claiming that doing so would be more in line with societal recognition of other First Amendment principles.[180]

The indirect harm recognized by advocates relates to the efficiency of such a website as a means of spreading white supremacist propaganda, which increases the risk of hate speech translating into harmful acts.[181] This argument is a narrower formulation of a closely related argument commonly made by advocates of hate speech regulation—the argument that hate speech can lead to violence.[182] Advocates tend to rely on studies of social psychologists to support these types of harm theories.[183] Of the social psychologists who have studied this link, Gordon W. Allport is widely recognized and commonly referenced by advocates for his work, having studied and identified five stages of the progression of racism: (1) antilocution (*i.e.*, hateful or racist speech); (2) avoidance; (3) discrimination; (4) physical attack; and (5) extermination.[184] Allport explains that activity on one level eases the transition to a more intense level.[185] Accordingly, the mere exist-

---

329 (7th Cir. 1985), *aff'd*, 475 U.S. 1001 (1986) ("Depictions of subordination tend to perpetuate subordination.").

[178] Delgado, *supra* note 171, at 146.

[179] *Id.* at 137.

[180] *See id.* at 145 (arguing that recognizing these harms would be more in line with societal recognition of the harm in separate but equal educational institutions and in the potential offensiveness of requiring one to display a state motto on their license plate); Lawrence, *supra* note 176, at 800 (explaining that promotion of self-expression and public discourse are two values protected by the First Amendment and implicated by hate speech); Matsuda, *supra* note 139, at 2377 ("As Professor Delgado has noted, the underlying first amendment values of self-fulfillment, knowledge, participation, and stable community recognized by first amendment theorists are sacrificed when hate speech is protected.").

[181] Romero, *supra* note 137, at 8.

[182] *See* Kretzmer, *supra* note 174, at 463 (discussing the potential of hate speech to lead to violence); Massey, *supra* note 128, at 155 (arguing that speech can incite enough hatred to lead to violence, and that suppressing speech is justified when done to protect individuals' private interests from invasion and protect society's interests in preserving order and peace within a community).

[183] Tsesis, *supra* note 14, at 869 (citing ALLPORT, *supra* note 172); *see also* Romero, *supra* note 137, at 12 n.30 (discussing arguments of social psychologists).

[184] *See* Kretzmer, *supra* note 174, at 463 (citing GORDON W. ALLPORT, THE NATURE OF PREJUDICE 285–339 (1954)) (discussing stages in progression of racism); Romero, *supra* note 137, at 12 n.30 (citing ALLPORT, *supra* note 172, at 142) (providing Allport as a reference for a "scholarly social science approach to understanding racial prejudice"); Tsesis, *supra* note 14, at 841 (citing ALLPORT, *supra* note 172) (discussing potential progression of racist speech to violence). Antilocution describes "hostile talk, verbal denigration and insult, and racial jokes" directed towards a group. RICHARD GROSS & NANCY KINNISON, PSYCHOLOGY FOR NURSES AND HEALTH PROFESSIONALS 146 (2d ed. 2014).

[185] Kretzmer, *supra* note 174, at 463 (citing ALLPORT, *supra* note 172).

ence of racist speech and hate speech generally makes the progression to the more violent stages easier.[186]

Advocates emphasize that other unique features of Internet communications provide a new context for hate speech that allows for a risk of potential harm greater than that which psychologists contribute to hate speech in general.[187] Advocates highlight the fact that the Internet facilitates development of communities of like-minded people, making it easier for those who plan to commit violence to connect.[188] This can also lead to normalization of violence within those communities, especially because counter-speech has proven less effective at combating online communications of hate speech.[189] Advocates often point to other nations for support, emphasizing the United States' position as an outlier and discussing the administrative issues that this position has caused for other nations attempting to combat online hate speech.[190] According to advocates, the global reach of online communications has led those who disseminate

---

[186] *Id.* Although Allport focuses on hate speech in general, scholars have relied on his findings in making arguments relating to this speech in an online context as well. *See, e.g.*, Tsesis, *supra* note 14, at 869 (citing ALLPORT, *supra* note 172).

[187] *See* Lidsky, *supra* note 151, at 149 ("[T]he actual or practical anonymity of many social media communications also fosters a sense of disinhibition in those contemplating violence, and the speed of communications allows incendiary speech to reach individual audience members at the point when they are most vulnerable to engaging in violent action."); Hammack, *supra* note 111, at 81 ("Now that the Internet has become an integral part of our culture, its ability to reach widespread audiences, rapid exchange of information, low cost of use, veil of anonymity, and constantly changing audience make threats posted on the Internet seem more dangerous than the same threats made in an off-line context.").

[188] *See* Lidsky, *supra* note 151, at 149 (explaining the potential of social media to foster the formation of "subcommunities of hate"); Hammack, *supra* note 111, at 82 ("Through email, discussion boards, and instant messaging, the Internet also facilitates the creation of networks of like-minded persons to help carry out threats").

[189] *See* Lidsky, *supra* note 151, at 149 (discussing the potential for community-building aspects of social media to foster or normalize violence); Hammack, *supra* note 111, at 81 (discussing the inability for counter-speech to combat harmful speech on the Internet due to "the widespread and transient nature of the Internet's audience," which makes it "virtually impossible to locate a discreet audience to refute objectionable speech").

[190] *See* Matsuda, *supra* note 139, at 2346–48 (contrasting the United States' position with other nations and emphasizing that it is in conflict with the international trend towards barring hate speech); Tsesis, *supra* note 14, at 863 ("These examples suggest that [the] United States['s] pure speech jurisprudence is anomalous and that it is generally accepted, by democracies like Canada and Germany, that preserving human rights supersedes the right of bigots to spread their venomous messages."); Webb, *supra* note 149, at 446–47 (describing the United States' role as a "safe haven for the promotion of hate speech" and noting that the United States' stance has undermined international efforts to combat hate speech). In response to arguments that divergence in positions on hate speech between the United States and other democracies may be justified by the unique history of those countries in combating discrimination and hate (e.g., Germany and its history with anti-Semitism), Alexander Tsesis presents a compelling argument: "[t]he history of racism in the United States, from Native American dislocation, to slavery, to Japanese internment, makes clear that here, as in other democracies, intolerance and persecution can exist in spite of the socially held ideal of equality." Tsesis, *supra* note 14, at 863.

hate speech to create domains in the United States, thus circumventing the laws of their own states.[191]

Even with the support of social psychologists and other nations, the difficulty in accurately and effectively conducting a study to demonstrate the degree of the link between violence and hate speech has made this argument vulnerable to attack by critics.[192] Critics attack the harm-based rationale, insisting that the harms contemplated by advocates are either not reliably measurable or are the subject of biased data or misconceptions in perception.[193] Critics place much weight on the relatively attenuated and somewhat uncertain connection between hate speech and violence, which makes it a common point of attack.[194] Advocates counter these attacks by pointing to individual cases where hate speech and violence are clearly linked.[195] They insist that, in light of the evidence that does exist in support of the connection between hate speech and social harms, the degree of potential harm at stake should trump any uncertainty in the precise degree of harm as the driving force for any decisions on whether to adopt regulations.[196]

---

[191] *See* Webb, *supra* note 149, at 446–47 (arguing that, "by permitting hate speech to flow freely within its borders, the United States undermines other nations' efforts to stop the promotion of hate speech"); *All Things Considered: Comparing Hate Speech Laws in the U.S. and Abroad*, NPR (Mar. 3, 2011), https://www.npr.org/2011/03/03/134239713/France-Isnt-The-Only-Country-To-Prohibit-Hate-Speech [https://perma.cc/3R3W-B8J4] (discussing concerns expressed by the director of the Yale Initiative for the Interdisciplinary Study of Anti-Semitism over the fact that hate groups have begun taking advantage of this by basing websites in the United States).

[192] *See* Bennett, *supra* note 131, at 476 (attacking harm-based arguments based on a weak causal connection); Strossen, *supra* note 127, at 250, 259 (same).

[193] *See* Bennett, *supra* note 131, at 499–500 (attacking the harm-based rationale for these reasons). In response to advocates who emphasize the psycho-emotional harms of hateful speech, John T. Bennett argues that using a psychological state as a metric of the impact of speech on psychological well-being is problematic due to an inability to precisely measure or gauge such a state on a scale. *Id.* at 486.

[194] *See, e.g.*, *id.* at 476 (attacking the harm-based argument based on uncertain and potentially-weak causal connection); Strossen, *supra* note 127, at 250, 259 (same). For example, Bennett insists that advocating for regulation based on social inequalities perceived to be the product of racism involves an "empirically unsound" line of reasoning. Bennett, *supra* note 131, at 476. He discusses various factors that might causally influence and contribute to social inequalities and explains that regulation would be misguided, at least to the extent these social and economic harms are influenced and caused by factors independent of hate speech. *Id.* Alternatively, Bennett posits that the harms are simply not severe enough to warrant regulation. *Id.* at 499–500. He suggests that, instead of promoting social justice, calls for regulation "entrench the status quo, which uplifts no one and is actually debilitating." *Id.* at 500.

[195] *See* Kretzmer, *supra* note 174, at 465 ("[W]hile general studies showing the connection between racist speech and the spread of racial discrimination or racist violence are almost impossible to execute, there is no lack of individual cases in which the connection between speech and violence has been quite clear."); Romero, *supra* note 137, at 9 (defending recognition of "a causal link between website hate speech and crimes committed by race mongers inspired by such propaganda" and providing Timothy McVeigh, the Oklahoma City bomber, as an example of an individual with direct ties to an anti-government, anti-minority movement).

[196] *See* Massey, *supra* note 128, at 155 (explaining ability of hate speech to incite violence, directed either towards the speaker or the targets, and declaring this sufficient justification for suppression of that speech); Tsesis, *supra* note 14, at 869 ("It is the paradox of any legal reform that remedies for social evils

3. Advocates' Proposed Solutions

Proponents of regulation offer a variety of approaches for amending current doctrine so as to encompass harmful speech online.[197] Many of these solutions involve changing the imminence standard so that hate speech can be properly evaluated under the incitement to imminent lawless action category of excepted speech.[198] These include proposals to modify the imminent standard only for specific instances, such as establishing a separate threshold of imminence appropriate for application only to advocacy of terroristic acts on the Internet.[199] Others are broader, advocating for an imminence standard applicable to all Internet speech.[200] For example, one commentator suggests modifying the standard for Internet speech by considering and balancing four factors in evaluating imminence: (1) imminence as evaluated from the listener's perspective, so as to capture the conduct that the incitement category of speech is meant to prevent; (2) the content of the speech, focusing on whether the speech, if spoken, would incite harm; (3) the likely audience; and (4) the nature of the issue involved, considering the value of its contribution to current debate.[201] Some have even suggested taking away the imminence requirement completely and imposing civil liability for speech that otherwise meets the incitement test under the *Brandenburg* standard, arguing that doing so does not hinder the ideals that the standard promotes.[202]

Approaches also include proposals to modify the standards under other categories of excepted speech to allow these categories to capture threatening speech that would otherwise escape categorization based on technicalities.[203] For

---

raise the possibility of new dilemmas. However, abstract uncertainties about potential evils should not constrain legislators from passing laws narrowly designed to curb expressions whose only object is to endanger . . . the less powerful.").

[197] *See, e.g.*, Crocco, *supra* note 153, at 483 (separate imminence threshold for serious advocacy of terroristic acts); Hammack, *supra* note 111, at 67 (refining approach under true threat standard to capture hateful and threatening speech online).

[198] *See, e.g.*, Crocco, *supra* note 153, at 483 (separate imminence threshold for serious advocacy of terroristic acts); Cronan, *supra* note 161, at 455 (Internet-style modification for incitement standard).

[199] *See, e.g.*, Crocco, *supra* note 153, at 483 (suggesting establishing a "threshold of imminence" that can apply Internet terrorism to justify regulating such speech).

[200] *See, e.g.*, Cronan, *supra* note 161, at 455 (proposing an Internet-compatible modification to the incitement standard).

[201] *Id.* at 455–57, 460.

[202] *See* Tiffany Komasara, *Planting the Seeds of Hatred: Why Imminence Should No Longer Be Required to Impose Liability on Internet Communications*, 29 CAP. U. L. REV. 835, 851, 853 (2002) (suggesting taking away the imminence requirement for incitement and imposing civil liability for speech on the Internet that otherwise meets the *Brandenburg* standard). The *Brandenburg* standard denies constitutional protection to speech that "is directed to inciting or producing imminent lawless action and is likely to incite or produce such action." Brandenburg v. Ohio, 395 U.S. 444, 447 (1969).

[203] *See* Hammack, *supra* note 111, at 67 (discussing the difficulty in applying true threat doctrine to online hate speech and arguing that courts should refine their approach to true threats to adapt to the changes brought about by the Internet). True threats are statements communicating an intent to commit

example, one commentator proposes modifying the intent standard under the true threat doctrine by considering the listener's objective fear and the speaker's subjective intent.[204] This allows the doctrine to capture threatening speech that fails under the true threat standard because the speaker did not threaten to perform the act herself and that fails under the incitement category because the lawless action is not technically "imminent."[205]

IV. EVALUATING TACTICAL, BUT NOT PRACTICAL ARGUMENTS

Online speech is unlike the traditional forms of speech in that it encompasses features and dimensions that are incompatible with First Amendment jurisprudence, but compatible with the underlying doctrinal goals of preserving and protecting a cohesive, democratic society.[206] Presented with this predicament, critics of regulation present two arguments: one in which they attack the degree of certainty in which one can prove the link between hate speech in violence, and another in which they adopt an unwavering, absolutist position, insisting on adhering to existing First Amendment doctrine without change or adjustment and claiming that destruction of our liberal democracy is at stake.[207] Not only is this a weak argument, but it also fails to recognize the realities of today's society.[208]

Section A of this Part analyzes and evaluates the tactics used by critics to detract and distract from the incompatibility of online speech and historic First

---

some violent, unlawful act on a specified group or individual, which are evaluated without regard to a speaker's actual, subjective intent. Virginia v. Black, 538 U.S. 343, 360 (2003).

[204] Hammack, *supra* note 111, at 67.

[205] *Id.*

[206] *See* Funk, *supra* note 132, at 580 (arguing that First Amendment doctrine does not contemplate, and thus does not directly address, the type of threatening and inciting speech communicated on the Internet); Lidsky, *supra* note 151, at 161 (discussing the problem with applying *Brandenburg* standard to social media and highlighting differences between audiences and speakers contemplated by *Brandenburg* test and those who use social media to express views); Hammack, *supra* note 111, at 66–67 (explaining that First Amendment jurisprudence relating to potentially threatening speech arose in the context of communications made in "fundamentally different media," where cases typically involved "remarks relayed to a very limited audience through pamphlets or at small rallies" and thus required a simpler analysis).

[207] *See* STROSSEN, *supra* note 134, at 155–56 (insisting that hate speech laws are ineffective in combating harms, might suppress protected speech, and "would gravely damage core principles that secure freedom of speech, equality, and democracy"); Bennett, *supra* note 131, at 478, 500 (arguing that the degree that speech fosters social inequalities is unclear and that a variety of causes, unrelated to racism, could be responsible); Crocco, *supra* note 153, at 457 ("Proponents of the Internet tend to be free expression absolutists and, not unlike other purists, fiercely guard the right to free speech under any circumstances.").

[208] *See* Thomas B. Nachbar, *Paradox and Structure: Relying on Government Regulation to Preserve the Internet's Unregulated Character*, 85 MINN. L. REV. 215, 216 (2000) ("First impulses about the Internet often turn out [] wrong because the Internet is so profoundly different from previous objects of regulation. And so it seems with the . . . many who have called for government to take a hands-off approach to the Internet, particularly in the area of speech regulation . . . .").

Amendment jurisprudence during attempts to argue against doctrinal change.[209] Section B argues that, in light of these transparent tactics and the weak arguments in favor of continued protection of hate speech, the United States must amend First Amendment doctrine to properly contemplate, consider, and protect against the realities of and interactions among cyberspace, socio-psychological influences of human behavior, and hate speech.[210]

A. A Classic Tactic: Distract, Accuse, and Distract Again

First Amendment absolutists who oppose restricting hate speech in ways that would disrupt or alter current First Amendment principles respond to advocates with three tactics: distraction, attack, and appeals to tradition and stagnancy.[211] Notably, none of these responses defend hate speech, in and of itself, as worthy of protection.[212] Instead, critics distract with a "parade-of-horribles"-type argument, in which they exclaim that restricting any speech will inevitably lead to erosion of the entire First Amendment doctrine and its protection.[213] The fallacy of this tactic becomes transparent when one looks at the judicially created origin and development of those categories of speech that are excluded from First Amendment doctrine.[214] The wavering of First Amendment protection for inciting speech during wartimes demonstrates this, as the level of protection granted to speech was contingent upon societal and political conditions, manifesting itself in the form of reduced protection for dissident speech during times

---

[209] *See infra* notes 211–227 and accompanying text.

[210] *See infra* notes 228–262 and accompanying text.

[211] *See, e.g.*, Bennett, *supra* note 131, at 521 (distracting with a "parade-of-horribles" argument by arguing that "hate speech regulation would empower lawmakers to barter away the right to free speech"); *id.* at 471 (attacking the harm-based rationale for restricting hate speech based on the assertion that social scientists have been associated with a liberal-leaning, institutional bias and alleging that it would be "constitutionally unsound to accept biased social science as a basis for restricting speech"); *id.* at 489, 491 (appealing to tradition by suggesting that the severity of racism inflicting American society does not warrant changing First Amendment doctrine to suppress hate speech and premising this on the fact that minorities have made significant progress towards true equality since the Civil Rights era).

[212] *See* Weintraub-Reiter, *supra* note 3, at 161 (acknowledging the higher costs and harms associated with hate speech on the Internet but defending the United States' protection of such speech based on the Constitution); Chemerinsky, *supra* note 139 (discussing harm of hate speech but advocating for free speech over regulation).

[213] *See, e.g.*, Strossen, *supra* note 127, at 250 ("Allowing speech to be curtailed on the speculative basis that it might indirectly lead to some possible harm sometime in the future would inevitably unravel free speech protection.").

[214] *See* Gilbert v. Minnesota, 254 U.S. 325, 332 (1920) (explaining that freedom of speech is not absolute but, rather, is subject to restrictions and limitations decided by the Court and explaining distinctions developed for such limitations in prior cases) (first citing Schenck v. United States, 249 U. S. 47, 52 (1919); then citing Frohwerk v. United States, 249 U. S. 204, 206 (1919)).

of war.[215] Failure to recognize that First Amendment protection has proven to be anything but absolute makes these arguments untenable.[216]

Critics of regulating hate speech attack any perceived weakness in the harm-based arguments that proponents champion.[217] Critics attack the purported link between violence and hate speech, calling it too attenuated or simply too dynamic in the factors contributing to it.[218] Most critics do not present any evidence to the contrary; they simply deny that the harms are definite enough or severe enough to warrant speech restrictions.[219] They distract from any links between hate speech and real-world violence that have been identified by ignoring examples of violent behavior spurred by hate speech online, honing in on the uncertainty inevitable in all scientific studies, and using the complex realities of human behavior and cyberspace to cast doubt on and distract from the evidence that does support a causal connection between Internet hate speech and violent acts.[220] Even the select few who do offer more legitimate support for their argu-

---

[215] Hudson, *supra* note 66, § 3:3; *see also Gilbert*, 254 U.S. at 338 (justifying curtailing of rights during times of war).

[216] *See* Crocco, *supra* note 153, at 457 ("Proponents of the Internet tend to be free expression absolutists and, not unlike other purists, fiercely guard the right to free speech under any circumstances."); *cf.* Gitlow v. New York, 268 U.S. 652, 667 (1925) (applying previous, less protective standard for incitement speech and emphasizing that the fact that a state may prohibit "utterances inimical to the public welfare, tending to corrupt public morals, incite to crime, or disturb the public peace, is not open to question") (first citing Robertson v. Baldwin, 165 U.S. 275, 281 (1897); then citing Patterson v. Colorado, 205 U.S. 454, 462 (1907); then citing Fox v. Washington, 236 U.S. 273, 277 (1915)); *see also* Warren v. United States, 183 F. 718, 721 (8th Cir. 1910) (explaining that the competency of Congress to restrain the rights of liberty and freedom of speech "in the interest of the general welfare, peace, and good order" is "beyond question," and as such, related legislation is consistently upheld by courts).

[217] *See, e.g.*, Bennett, *supra* note 131, at 500 (basing his attack on speech-based harms on difficulties in reliably measuring social harm and an allegation that there is no reason to believe speech-based harm is as severe as advocates insist). In response to advocates who emphasize the psycho-emotional harms of hateful speech, Bennett argues that using a psychological state to gauge the impact of speech on psychological well-being is problematic due to an inability to precisely measure such a state on a scale. *Id.* at 487.

[218] *See, e.g.*, STROSSEN, *supra* note 134, at 155–56 (insisting that hate speech laws are ineffective in combating harms, might suppress protected speech, and "would gravely damage core principles that secure freedom of speech, equality, and democracy"); Bennett, *supra* note 131, at 478, 500 (arguing that the degree that speech fosters social inequalities is unclear and that a variety of causes, unrelated to racism, could be responsible); *see also* Bennett, *supra* note 131, at 491 (defending against harms of hate speech by arguing that "[e]xpressions of overt racism in modern America have undeniably decreased following the Civil Rights era").

[219] *See, e.g.*, Bennett, *supra* note 131, at 500 ("In summary, there is no reason to believe that American minorities are facing harms of racism and discrimination *to the degree* posited by Matsuda, Delgado, and other speech regulation advocates. The core premises of hate speech regulation could be erroneous."); *accord id.* at 489 (stating that "[r]acism may at times be misperceived" in support of his argument that psychological and emotional harm should not be relied upon to gauge harm caused by speech). *But see* STROSSEN, *supra* note 134, at 151–52 (discussing psychological studies suggesting that shielding individuals from stress-inducing speech may reduce mental health).

[220] *See, e.g.*, STROSSEN, *supra* note 134, at 151–52 (emphasizing the benefits of short-term stress on mental health and crediting hate speech as a source of such benefits); Bennett, *supra* note 131, at 478 (basing his arguments on the premise that societal harms flowing from racism and hate speech are not as

ments appear to resort to such tactics, whether intentional or not, as a veil for the weak support for their positions.[221] One prominent scholar, for example, relies on the findings of two international, observation and experience-based studies for her rejection of the merits of the causal connection between hate speech and harm.[222] While these studies do offer legitimate support, they were conducted in the early 1990s and resulted in no definitive conclusions regarding such connection, apart from the inability to identify one.[223] It is difficult to imagine that such studies would conclude similarly if conducted after the advent of social media.[224]

Finally, critics distract again by pointing to countries with hate speech laws that have been abused and suggest that hate speech regulations are susceptible to abuse in the United States.[225] Interestingly enough, unlike they did in their anal-

---

bad as academic researchers suggest and arguing that the potential number of causes that contribute to societal harms favors resisting regulation). For example, Bennett discusses the position of hate speech regulation advocate Professor Delgado, who analyzed social psychology research studies that indicate the harmful impact of racist speech on minorities. *Id.* at 488. Rather than presenting contradictory evidence, Bennett rejects the study as outdated and argues that, because the study was conducted in 1968, it should not be relied upon as support for the harms of racist speech in modern times. *See id.* ("Delgado's citation for that claim was a study published in 1968. Surely the impact of racism in America has changed somewhat during the intervening years . . . . Those committed to the tradition of free speech may want answers to these questions before consenting to surrendering their First Amendment rights.").

[221] *See, e.g.*, Strossen, *supra* note 127, at 258–59 (relying upon findings of studies conducted prior to the rise of the Internet).

[222] *See id.* (discussing studies).

[223] *See id.* (discussing studies).

[224] *See id.* (discussing studies from 1991 and 1992); *see also* STROSSEN, *supra* note 134, at 136–37 (relying on Human Rights Watch studies to argue against hate speech restrictions). *See generally* COMMON SENSE MEDIA, *Percentage of Teenagers in the United States Who Have Encountered Hate Speech on Social Media Platforms as of April 2018, by Type*, STATISTA, https://www.statista.com/statistics/945392/teenagers-who-encounter-hate-speech-online-social-media-usa/ [https://perma.cc/93WB-AKQF] (2018 survey of 1,141 U.S. teenagers, ages thirteen to seventeen, showing that fifty-two percent reported having encountered hate speech on social media often or sometimes); Felix Richter, *The Rise of Social Networking in the United States*, STATISTA (2013), https://www.statista.com/chart/913/the-rise-of-social-networking-in-the-united-states/ [https://perma.cc/VEH3-KG95] (detailing rise of social media networking in the United States from 2005 to 2013); Patrick Wagner, *By 2021 More Than 1/3 of the Globe Will Be on Social Media*, STATISTA (2018), https://www.statista.com/chart/15355/social-media-users/ [https://perma.cc/YY5X-KWED] (providing statistics on global social media use, spanning from 2010 to 2017, and predicting outcome for the years 2018 to 2021 based on upward trend). Acknowledging the potential impact of social media, Strossen points to another Human Rights Watch study conducted in 2016, which focused on hate speech regulations in India and reached similar results. *See* STROSSEN, *supra* note 134, at 83 ("A quarter-century later, Human Rights Watch reached a similar conclusion in its report on the enforcement of 'hate speech' laws in India."). The narrow scope of this report, focused only on one country, makes this contention unpersuasive. *See id.* (justifying reliance on older Human Rights Watch study with results in 2016 study of hate speech in India).

[225] *See, e.g.*, STROSSEN, *supra* note 134, at 81 ("Given the pervasiveness of individual and institutional bias, the government is likely to enforce 'hate speech' laws, as it has other laws, to the disadvantage of disempowered and marginalized groups. Indeed, laws censoring 'hate speech' have predictably been enforced against those who lack political power . . . ."); Strossen, *supra* note 127, at 258 (discussing selective enforcement of laws banning hate speech in South Africa).

ysis of the potential harms of hate speech, critics place little emphasis on the abundance of other causal factors that might contribute to this abuse, such as the political, social, or economic climates in those countries.[226] In contrast, proponents of hate speech regulation list countries with similar democratic institutions and values in place, such as Germany and Canada, which have successfully promulgated those regulations without destroying the fabric of their liberal democracies, thus revealing the transparency of these tactics.[227]

B. A Rational Rethinking for the Real World

Much of the scholarship surrounding the First Amendment's protection of hate speech on the Internet emerged prior to the advent of social media.[228] Consideration of the impact of social media that social psychologists have more recently studied, however, not only lends support to the arguments advanced by proponents of restricting hate speech, but also further demonstrates the weaknesses of the arguments advanced by critics of regulation.[229] For example, a common argument against regulation appeals to the power of counter-speech, suggesting that a plausible solution to the problem of hate speech is to fight it with counter-speech, which avoids violating existing First Amendment principles.[230] This argument, however, presumes an effectiveness of counter-speech

---

[226] *See, e.g.*, Strossen, *supra* note 127, at 258–59 (citing STRIKING A BALANCE: HATE SPEECH, FREE SPEECH, AND NON-DISCRIMINATION, *supra* note 127) (discussing failure of hate speech laws in other countries and omitting discussion of other factors that may distinguish the viability of such laws in the United States).

[227] *See* Matsuda, *supra* note 139, at 2346–47 ("[T]he existing domestic law of several nations—including states that accept the western notion of freedom of expression—has outlawed certain forms of racist speech."); Timofeeva, *supra* note 25, at 254 (contrasting the United States' approach with that of Germany and explaining that both nations share commitment to freedom of speech and traditional liberalism tenets); Webb, *supra* note 149, at 446–47 ("Today, nearly every nation across the globe regulates hate speech in some way to promote human dignity and protect minorities from verbal persecution. The United States, however, rests in the minority, and it remains the only country to expressly protect it.").

[228] *See* Romero, *supra* note 207, at 3 (discussing inability for First Amendment jurisprudence to combat the spread of white supremacist websites and making no reference to social media); Weintraub-Reiter, *supra* note 3, at 157–58 (discussing modes of communication and information retrieval offered by the Internet but omitting social media from list).

[229] *Compare* STROSSEN, *supra* note 134, at 182 (defending counter-speech as a remedy for hate speech), *and* Franklyn Haiman, *The Remedy Is More Speech*, AM. PROSPECT (1991), https://prospect.org/article/remedy-more-speech [https://perma.cc/GE4B-DKSU] (same), *with* Thai, *supra* note 56, at 310 (suggesting the inadequacy of counter-speech as a remedy to combat harmful speech and basing this on the inability "to reach the highly polarized echo chambers of social media"), *and* Hammack, *supra* note 111, at 81 (discussing the inability for counter-speech to combat harmful speech on the Internet due to the nature of audience, making it "virtually impossible to locate a discreet audience to refute objectionable speech").

[230] *See, e.g.*, David L. Hudson, Jr. & Mahad Ghani, *Hate Speech Online*, FREEDOM F. INST. (Sept. 18, 2017), https://www.freedomforuminstitute.org/first-amendment-center/topics/freedom-of-speech-2/internet-first-amendment/hate-speech-online/ [https://perma.cc/3URR-B6TF] (discussing the chair of the Anti-Defamation League's Internet Task Force's view that "[c]ounter-speech is a potent weapon" in combating hate speech).

that the echo chambers effect impedes by filtering out what would be counter-speech and exposing a user solely to content supportive of his or her own views.[231] Social media algorithms designed to personalize user content feeds contribute to further this effect, allowing for amplification of the polarizing impact of confirmation bias and the creation of echo chambers.[232] These processes all interact to contribute to polarization on both an individual and societal level and, most importantly, increase the potential of hate speech to translate into actual violence.[233] Filtering out a countervailing view distorts the prevalence of the remaining view as perceived by both the speaker and listener of hate speech.[234] To victims of hate speech, this amplifies the impact of the hate speech in terms of subordination and denigration felt.[235] To the speakers, the absence of counter-speech encourages feelings of validation in their views and actions, which can lead to adoption of more radical stances or even violence.[236] Thus, although not "imminent" enough to amount to create a clear and present danger under the formal *Brandenburg* test, the severity, directness, and ability of this harm to incite becomes increasingly clear.[237]

Re-evaluating domestic terror attacks and past incidents of violence, in light of what is known about this phenomenon, demonstrates the reality behind

---

[231] *Compare* Weintraub-Reiter, *supra* note 3, at 162 (defending counter-speech as a weapon and attributing the creation of websites devoted to the history of the Holocaust to "anger and activism" incited by websites that deny the Holocaust), *with* Freilich, *supra* note 53, at 692–93 (discussing the improbability of counter-speech successfully overcoming "radicalization echo chamber[s]" created by terrorist groups), *and* Hammack, *supra* note 111, at 82 ("[I]f an anti-Semitic web site publishes falsehoods slandering Jews, visitors to that site would be unlikely to visit a Jewish organization's web site refuting the anti-Semitic speech."). The effectiveness of counter speech is significantly hindered by human nature. *See* Am. Booksellers Ass'n, Inc. v. Hudnut, 771 F.2d 323, 328–29 (7th Cir. 1985), *aff'd*, 475 U.S. 1001 (1986) ("People often act in accordance with the images and patterns . . . around them. People raised in a religion tend to accept the tenets of that religion, often without independent examination . . . . Even the truth has little chance unless a statement fits within the framework of beliefs that may never have been" rationally studied).

[232] *See* Andorfer, *supra* note 33, at 1414–15 (describing social media algorithms); Farag, *supra* note 49, at 863 (detailing ISIS's strategic use of echo chambers and online communities to distort perceptions to promote radicalization).

[233] *See* SUNSTEIN, *supra* note 31, at 131 (discussing the echo chambers effect and role of social media in contributing to polarization and in influencing perception and behavior); Lidsky, *supra* note 151, at 149 (discussing the characteristics of online communication and an increased potential for violence); Hammack, *supra* note 111, at 82 (same).

[234] *See* Farag, *supra* note 49, at 863 (citing Neumann, *supra* note 53, at 435–36) (discussing this process in the context of terrorist-created radicalization "echo chambers").

[235] *See* Delgado, *supra* note 171, at 137 (discussing psychological responses to racial stigmatization).

[236] *See* Hammack, *supra* note 111, at 94 (discussing the ability of the Internet to encourage the formation of social groups online, where users "can encourage and facilitate threatening behavior").

[237] *See* Cronan, *supra* note 161, at 456 (arguing that the "goals of preventing the undesirable consequences of incitement" cannot be attained on the Internet without altering the interpretation of the imminence requirement); Delgado, *supra* note 171, at 137 (discussing harms).

the manifestation of hate speech into real-life violence.[238] In 2015, for instance, Dylann Roof murdered nine individuals at a historical black church in South Carolina; his online rantings suggest this phenomenon's involvement.[239] On his website, Roof complained that "[w]e have no skinheads, no real KKK, no one doing anything but talking on the internet."[240] This indicates that Roof noticed a distortion between the amount of individuals he "talk[ed] on the internet" with and the amount of action being taken in the real world, and it is very possible that this distortion contributed to his determination to act and his validation of those acts.[241] Although this is only one example, many domestic terrorists have similarly engaged in online hate speech.[242] As social media becomes an increasingly significant part of our society's daily interactions, the potential for this phenomenon to repeat itself is limitless and daunting.[243]

In fact, concerns about an increase in incidents of domestic terrorist attacks and increases in the amount of hate speech online have recently emerged, closely tracking the rise of social media.[244] Rather than brushing off these concerns as too speculative or demanding more definitive research studies, those who advocate against hate speech regulations should consider what is happening in the real world.[245] Knowledge about the impacts of social media on human behavior

---

[238] *See* Farag, *supra* note 49, at 844, 863 (discussing the rising number of Americans prosecuted for charges relating to the Islamic State and noting the role of social media, the echo chamber effect, and the Internet in those cases); Goldman, *supra* note 26 ("After virtually every mass shooting, every high-profile hate crime over the past decade, the story played out much the same: All the warning signs were on full display on social media.").

[239] Sanchez & Payne, *supra* note 4; *see* Robles, *supra* note 5 (explaining the content behind Dylann Roof's online posts).

[240] Robles, *supra* note 5.

[241] *See id.* (reporting Dylann Roof's childhood friend's statement that "[t]his whole racist thing came into him within the past five years").

[242] *See generally* JEROME P. BJELOPERA, CONG. RESEARCH SERV., R44921, DOMESTIC TERRORISM: AN OVERVIEW 48–49 (2017) (providing a comprehensive overview of the history of domestic terrorism in the United States).

[243] *See* Weintraub-Reiter, *supra* note 3, at 161 ("Speech on the Internet, specifically hate speech, now has a wider audience than other media, and, therefore, the societal costs will be higher."); *see also* Matsuda, *supra* note 139, at 2360 (arguing that we should accept the collective and international rejection of racial supremacy as a "shared historical legacy of the world community," which has been recognized as harmful and dangerous to its victims and society).

[244] *See* Paul K. McMasters, *Must a Civil Society Be a Censored Society?*, 26 HUM. RTS. 8, 9 (1999) (discussing the debate on both sides, commenting on the problem of the rise of the Internet and its impact on hate speech, and explaining that most Americans want to do something about the hate); IRISH HUMAN RIGHTS & EQUAL. COMM'N, *Press Release: Human Rights and Equality Commission Challenges Rise of Hate Speech Online* (Nov. 28, 2018), https://www.ihrec.ie/human-rights-and-equality-commission-challenges-rise-of-hate-speech-online/ [https://perma.cc/579D-Y5L8] (discussing hate speech's increasingly prominent presence online and urging for Ireland to take on the role of an international leader in fighting the spread of hate speech on the Internet).

[245] *See* McMasters, *supra* note 244, at 8, 9 (discussing the rise of hate speech on Internet and problematic consequences); Strossen, *supra* note 127, at 250 (warning that proscribing speech on a speculative basis of future harm would lead to unravelling of free speech protection); IRISH HUMAN RIGHTS & EQUAL. COMM'N, *supra* note 244 (discussing hate speech's increasingly prominent presence online and

and psychology may be limited, but what is known is significant.[246] Moreover, with hate speech on the rise and social media continuing to develop and make way for novel modes of communication, the potential for future harm is at its peak.[247] Rejecting that harm as speculative, or denying the existence of any tangible, direct harms beyond the psychological injuries to hate speech victims, is not only mistaken but also deeply problematic.[248] This is even more so where, as is the case here, those who reject that harm as insufficient enough to warrant denial of First Amendment protection present arguments that do not hold weight.[249]

As social media networks grow in prominence and global usage, failure to address the shortcomings of American constitutional doctrine will only become more problematic, polarizing, and detrimental to society.[250] As a nation, we must accept the failings of our current doctrine and embrace the fact that Internet speech is vastly different than the forms of speech contemplated during ratification of the Constitution and throughout much of First Amendment doctrinal development.[251] Rethinking the imminence standard under the incitement to imminent lawless ac-

---

encouraging action to fight the spread). For example, in Sri Lanka, after an attack that left more than 290 dead, the government shut down various social media platforms—a decision made "out of fear that misinformation about the attacks and hate speech could spread, provoking more violence." Max Fisher, *Sri Lanka Blocks Social Media, Fearing More Violence*, N.Y. TIMES (Apr. 21, 2019), https://www.nytimes.com/2019/04/21/world/asia/sri-lanka-social-media.html [https://perma.cc/6CJS-KSYU].

[246] *See* Timofeeva, *supra* note 25, at 254–55 (contrasting the United States' approach with that of Germany and calling differences "particularly disturbing" given that regulatory efforts of one nation might be hindered by another nation's stance).

[247] *See* Weintraub-Reiter, *supra* note 3, at 146 (discussing the fact that the many modes of communication available through cyberspace have been used as tools for the perpetuation of hate speech and omitting any discussion of later-arising social media). For a discussion on the rise of hate speech on social media, see Luiz Valério P. Trindade, DISCOVER SOC'Y, *On the Frontline: The Rise of Hate Speech and Racism on Social Media* (Sept. 4, 2018), https://discoversociety.org/2018/09/04/on-the-frontline-the-rise-of-hate-speech-and-racism-on-social-media/ [https://perma.cc/7YZN-RA5N].

[248] *See* Strossen, *supra* note 127, at 250 ("Allowing speech to be curtailed on the speculative basis that it might indirectly lead to some possible harm sometime in the future would inevitably unravel free speech protection."); *see also* Timofeeva, *supra* note 25, at 254 ("It is widely recognized that hate propaganda harms society as a whole.").

[249] *See, e.g.*, Weintraub-Reiter, *supra* note 3, at 162 (discussing the creation of websites devoted to the history of the Holocaust as a result of the "anger and activism" impelled by websites that deny the Holocaust); *see also* Freilich, *supra* note 53, at 692–93 (discussing the improbability of counter-speech successfully overcoming "radicalization echo chamber[s]" created by terrorist groups).

[250] *See* Wagner, *supra* note 225 (providing statistics on global social media use, spanning from 2010 to 2017, and predicting outcome for the years 2018 to 2021 based on upward trend).

[251] *See* Funk, *supra* note 132, at 580 (arguing that First Amendment doctrine does not contemplate, and thus does not directly address, the type of threatening and inciting speech communicated on the Internet); Lidsky, *supra* note 151, at 148–50, 160 (discussing the problem with applying *Brandenburg* standard to social media and highlighting differences between audiences and speakers contemplated by *Brandenburg* test and those who use social media to express views); Hammack, *supra* note 111, at 66 (explaining that First Amendment jurisprudence relating to potentially threatening speech arose in the context of communications made in "fundamentally different media," where cases typically involved "remarks relayed to a very limited audience through pamphlets or at small rallies" and thus required a simpler analysis).

tion category of speech provides a good, if not necessary, starting place.[252] While the requirement of imminence traditionally provides a safeguard against frivolous over-censorship of speech based on consequences that are too far-removed, cyberspace challenges previous notions of time and space and contemplates less transparent behavioral responses.[253] Hiding behind a computer screen, speakers of potentially harmful speech can now incite violence or spread hatred from another part of the world and do not have reason to fear physical harm or any other consequence of physical presence stemming from incitement of a crowd or individual.[254] With this possibility for remote and shielded incitement, and the widespread availability and efficiency of social media for disseminating speech, the costs of participating in and propelling hate speech have lowered.[255] The severity of the associated harms, however, have risen.[256]

The rising prevalence of online hate speech merits more than simply tweaking current doctrine as a compromise with tradition.[257] It is important to remember the judicially created origin of these categories and, within these categories, the fluctuation of the standards required to censor such speech during times of peace and times of war.[258] The current First Amendment jurisprudence is a result of judi-

---

[252] *See* Hudson, *supra* note 66, § 3:3 (explaining a pattern of greater government restriction on speech during times of war and lower restriction during times of peace); *see also Gilbert*, 254 U.S. at 338 ("There are times when those charged with the responsibility of Government, faced with clear and present danger, may conclude that suppression of divergent opinion is imperative; because the emergency does not permit reliance upon the lower conquest of error by truth. And in such emergencies the power to suppress exists.").

[253] *See* Timofeeva, *supra* note 25, at 253–54 (describing unique characteristics of the Internet, including the variety of communicative options in terms of parties involved and audience number, lack of inherent restrictions on size or resources, and its provision of globalism and anonymity); Hammack, *supra* note 111, at 67 ("The unique characteristics of the Internet blur the distinction between threats and incitement by allowing speakers to threaten by incitement . . . .").

[254] *See* Lidsky, *supra* note 151, at 149 ("[T]he actual or practical anonymity of many social media communications also fosters a sense of disinhibition in those contemplating violence, and the speed of communications allows incendiary speech to reach individual audience members at the point when they are most vulnerable to engaging in violent action.").

[255] *See* Timofeeva, *supra* note 25, at 253–54 (describing unique characteristics of Internet, including widespread audience, lack of inherent restrictions on size or resources, and its provision of globalism and anonymity); Hammack, *supra* note 111, at 81 (explaining that the Internet's "ability to reach widespread audiences, rapid exchange of information, low cost of use, veil of anonymity, and constantly changing audience make threats posted on the Internet seem more dangerous than the same threats made in an offline context").

[256] *See* Weintraub-Reiter, *supra* note 3, at 146 ("Cyberspace has been and continues to be used to perpetuate hate speech.").

[257] *Cf.* Timofeeva, *supra* note 25, at 254 ("In spite of many new communicative and technical options of the Internet, both the United States and Germany attempt to fit this new media into their old free speech standards.").

[258] *See* Hudson, *supra* note 66, § 3:3 (describing a pattern of greater protection of speech during times of peace and lesser during times of war); *see also Gilbert*, 254 U.S. at 338 ("There are times when those charged with the responsibility of Government, faced with clear and present danger, may conclude that suppression of divergent opinion is imperative; because the emergency does not permit reliance upon the lower conquest of error by truth. And in such emergencies the power to suppress exists.").

cial policymaking in response to societal needs.[259] Regulating online hate speech would simply be a continuation of this traditional response.[260] As an outlier in its stance on hate speech, the United States has the unique opportunity to look to other nations' hate speech regulations, their implementation, and their impact in real-time.[261] Failure to take advantage of this would be sophomoric and indefensible.[262]

CONCLUSION

Because hate speech does not squarely fall within any of the categories excluded from First Amendment protection, the United States is an outlier in that, unlike most nations, it protects this hate speech. The inability of existing First Amendment doctrine to combat hate speech has led to criticism and calls for reform by those who believe the societal harms of hate speech warrant amending existing doctrine in favor of banning this speech. These calls garner support from social psychologists studying the psychological impacts of social media on behavior and psychologists who have studied the psychology behind hate speech, as these studies indicate that the harms that online hate speech present extend far beyond dignitary harms to victims, reaching society as a whole by promoting violence and disorder. Faced with this overwhelming reality, those who advocate against banning hate speech respond with tactics designed to distract from the reality behind these contentions, presenting a "parade-of-horribles" argument and attacking any perceived weakness in arguments posed by the other side. Revealing the fallacies behind these tactics demonstrates the need to amend First Amendment doctrine so that it can properly combat, control, and contemplate the power of hate speech transmitted through social media communications. Doing so requires more than a forced doctrinal amendment, and would be consistent with the judicial policy-based development of present First Amendment doctrine. With so many other nations leading the way to soften the risk, it is time for the United States to follow.

LAUREN E. BEAUSOLEIL

---

[259] *See Gilbert*, 254 U.S. at 338 (rationalizing the government's ability to respond to "clear and present danger" with suppression of speech); Hudson, *supra* note 66, § 3:3 (describing a pattern of greater protection of speech during times of peace and lesser during times of war).

[260] *See Gilbert*, 254 U.S. at 338 (rationalizing governmental suppression of speech in response to "clear and present danger"); Hudson, *supra* note 66, § 3:3 (observing reduced protection of speech during times of war).

[261] *See* Webb, *supra* note 149, at 446 (describing United States' role as a "safe haven for the promotion of hate speech"); Gonzalez, *supra* note 13 (discussing law governing hate speech in the United States and comparing with Poland, France, Germany, and the United Kingdom). For a comparative overview of various nations' stances on hate speech, see EMORE, AN OVERVIEW ON HATE CRIME AND HATE SPEECH IN 9 EU COUNTRIES, *supra* note 14, at 8.

[262] *See* Timofeeva, *supra* note 25, at 254–55 (contrasting the United States' approach with that of Germany and calling differences "particularly disturbing" given that regulatory efforts of one nation might be hindered by another nation's stance); Webb, *supra* note 149, at 446–47 (describing United States' role as a "safe haven for the promotion of hate speech" and noting that the United States' stance has undermined efforts to combat hate speech).